AFIT/GSM/LAS/96S-6

CALIBRATION OF THE SOFTCOST-R SOFTWARE COST MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

THESIS

Steven V. Southwell, B.S.
Captain, USAF

AFIT/GSM/LAS/96S-6

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

AFIT/GSM/LAS/96S-6

CALIBRATION OF THE SOFTCOST-R SOFTWARE COST MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

THESIS

Presented to the Faculty of the Graduate School of Logistics
and Acquisition Management of the Air Force Institute of Technology
Air University
Air Education and Training Command
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Logistics Management

Steven V. Southwell, B.S.
Captain, USAF

September 1996

Preface

This thesis effort determined the uncalibrated and calibrated accuracy of the SoftCost-R, Version 8.4, software cost estimation model's effort predictions for software development Computer Software Configuration Items (CSCIs) in the Space and Missile Systems Center (SMC) Software Database, Version 2.1 (SMC SWDB). It also determined what improvement in accuracy was achieved by the calibration.

I appreciate the support and assistance provided by my thesis advisor, Professor Daniel Ferens, and my reader, Dr. David Christensen. I would like to thank Ms. Sherry Stukes of MCR and Ms. Shirley Tinkler of SMC/FMC for providing the SMC SWDB, assisting with its use, and for comments to this thesis while it was in work. I would like to thank Mr. Anthony Collins of Resource Calculations, Inc. (RCI) for providing copies of the models and manuals for the SoftCost family of models (SoftCost-R, SoftCost-Ada, and SoftCost-OO) and for his willingness to provide answers and insight into the SoftCost-R model. I would also like to express my deepest gratitude to Mr. Joe Bruscino, software support programmer for SoftCost-R, for his assistance in determining the complete equations with which the model calculates software effort and for helping me figure out the parts of the documentation that were missing or incorrect.

Most importantly, I would also like to thank my wife, Penny, and children, Jonathan, Michael, and Sarah, for their support, patience and understanding. They are what makes it all worthwhile.

Steven V. Southwell

Table of Contents

Page

List of Figures

Figure Page

List of Tables

List of Equations

Abstract

The rising number and importance of Department of Defense software developments, when combined with declining defense budgets, has resulted in a critical need to accurately plan and manage software development costs and schedules. Unfortunately, the increasing size, complexity, and diversity of these software developments has made accurate estimating problematic. Uncalibrated software cost models have not generally produced reliable results due to generic default parameters and improper usage. The default parameters cannot hope to accurately represent and predict the wide variability of software efforts to which the models are being applied. However, some of the models have achieved improved accuracy by calibration from their generic default parameters to new parameter values based on specific characteristics of the development efforts being estimated. This research effort focused on the calibration of SoftCost-R, Version 8.4, to specific stratified data sets obtained from the Space and Missile Systems Center (SMC) Software Database, Version 2.1, (SWDB). The accuracy of the new calibrated inputs was verified through comparisons between the calibrated and default estimates and the actual cost data. Statistical methods used to make the comparisons included magnitude of relative error (MRE), mean magnitude of relative error (MMRE), root mean square (RMS), relative root mean square (RRMS), and prediction level Pred (k/n) or percentage of estimates within (100 * k/n)% of the actual costs. The new calibrated parameters resulted in more accurate effort estimates and the calibration method appeared to be valid. However, the accuracy improvement was neither complete nor all encompassing. That is, the calibrated goodness of fit did not meet Conte's criteria of MMRE $\leq$ 25%, RRMS $\leq$ 25%, or Pred (0.25) $\geq$ 75%, and not all of the data sets achieved significant accuracy improvement due to the calibration. This result

agrees with previous studies and emphasizes the need for complete, accurate, and homogeneous data.

CALIBRATION OF THE SOFTCOST-R SOFTWARE COST MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

I. Introduction

Overview

"Software is critical to the operation of our satellites, ships, submarines, aircraft, tanks, missiles, command and control systems and intelligence systems" (Lieblein, 1986:734). This statement is even more true today than when it was originally made. Due to the increasing amount and importance of software in Department of Defense (DoD) weapon systems, ever larger portions of system acquisition and support budgets are software related. When this trend is combined with decreasing DoD budgets, the result is a critical need to accurately plan future software development costs and schedules without jeopardizing mission critical software capabilities (Thibodeau, 1991:1-2; Pacheco, 1987:2). Unfortunately, the increasing size, complexity, and diversity of these software developments have made accurate estimating difficult.

General Issue

One method to plan for future software development costs is to use computerized, parametric software cost and schedule estimating models. These models utilize equations with input and internal parameters that were from past software development projects. The default values for the parameters, when combined with user inputs for software size, complexity, developer skill, and other factors, result in a forecast (estimate) for software cost and schedule. However, estimates obtained using the default parameter values

embedded in these models are usually inaccurate, and they tend to underestimate the cost, size, and schedule of future projects (Boehm, 1981:320-321, 330-333, 342; Brooks, 1975:14-16).

A recent case study performed at the Air Force Institute of Technology indicated that software cost models, have not produced reliable results (Ferens and Christensen, 1995:1). However, some of the models have achieved some accuracy improvement due to calibration from their generic default parameters to new parameter values based on specific characteristics of the development effort being estimated (Ferens and Christensen, 1995:15). These new calibrated inputs will hopefully result in more accurate estimates for similar future software development projects.

Specific Issue

The Air Force Space and Missile Systems Center (SMC) in Los Angeles has a Software Database (SWDB) of over two thousand software development efforts for which they desire to have several software estimating models calibrated to a level of specificity that will enable improved estimates for future software development projects. The SoftCost-R (Software Costing - Real-time) model, distributed by Resource Calculations, Inc. (RCI), Englewood CO, is one of the models that SMC desires to have calibrated.

Research Objective

By calibrating the productivity ratio factor, also known as the thousands of source lines of executable code productivity multiplier (KSLECPM), and the productivity adjustment constant, also known as the average work force factor (AWFFAC), for specific application types, this research aims to improve the fit between SoftCost-R, Version 8.4, and the SMC SWDB, Version 2.1 (Resource Calculations, Inc., undated:1). The following statistical tests shall be used to assess the goodness of fit: Magnitude of

Relative Error (MRE), Mean Magnitude of Relative Error (MMRE), Root Mean Square (RMS), Relative Root Mean Square (RRMS), and Prediction Level (Pred (0.25)) (Ferens and Christensen, 1995:8).

The research questions to be answered include:

1. What is the uncalibrated accuracy of the SoftCost-R model when estimating efforts in the SMC SWDB?
2. Can SoftCost-R be calibrated to subsets of the SMC SWDB?
3. What is the accuracy of the SoftCost-R model after it has been calibrated to the SMC SWDB?
4. What improvement was achieved due to calibration?

Scope of Research

The scope of this research effort is limited to calibration and validation of new development effort parameters derived from the SMC SWDB, Version 2.1. It does not include schedule, risk analysis, work allocation, or support effort parameters. Also, this research will not be evaluating, calibrating, or using the Constructive Cost Model (COCOMO) submodel (COCOMO-R) within SoftCost-R, Version 8.4, since the Revised Enhanced Version of the Intermediate COCOMO (REVIC) has already been calibrated to the SMC SWDB (Weber, 1995). The ability to use these new development effort calibration parameters in other environments is an area for future research; it is believed that new calibration parameters would need to be developed based on a database of projects for the desired environment.

Thesis Overview

This research will use the SMC SWDB, Version 2.1 to calibrate SoftCost-R, Version 8.4. Chapter II, Literature Review, reviews research efforts and documentation in the areas of software costs, software cost estimation, software cost model calibration, the

SMC SWDB, and the theory underlying SoftCost-R calibration. Chapter III, Methodology, describes the SMC SWDB, how the SMC SWDB records were chosen and stratified, how SoftCost-R was calibrated and validated, and how the estimating accuracy was assessed. Chapter IV, Findings, presents the results of the calibration, validation, and accuracy assessment. Finally, Chapter V, Conclusions and Recommendations, contains conclusions based on the findings and recommendations for further research.

Appendix A is a glossary of acronyms and technical terms. Appendix B is a correspondence matrix between the SMC SWDB and SoftCost-R. Appendix C contains calibration data and statistics. Appendix D contains validation data and statistics. Appendix E includes the validation Wilcoxon signed-rank test data.

II. Literature Review

Overview

This chapter is a review of research and discussions relevant to software costs, software cost estimating, software cost model calibration, the SMC SWDB, and the theoretical basis for calibrating SoftCost-R. It is intended to provide the reader with a brief synopsis of the basics of software cost estimation, the current status of the software cost estimation field, the SMC SWDB, and the SoftCost-R model.

Software Costs

Ever larger portions of system acquisition and support budgets for DoD weapon systems are software related. According to Wellman, software costs "now amount to about 90% of the total cost to the end user over the life cycle of the software" (Wellman, 1992:30). However, it is believed that Wellman was referring only to the software and computer hardware costs, not the entire system costs. DoD software life cycle costs in 1990 were around $34 billion (Marsh, 1990:62). Also, "A General Accounting Office study pointed out that more than 50 percent of the software systems studied had significant cost overruns; more than 60 percent had serious schedule slippages" (Putnam and Myers, 1992:9). Wellman also reported that 65% of software projects exceed their initial budget by more than 25% (Wellman, 1992:11). These inaccurate estimates can result in serious cost overruns or cancellation of a project when initial estimates are too low and missed opportunities when initial estimates are too high and a project is not undertaken (Wellman, 1992:13; Jones, 1994:155).

Cost Estimating Models

According to Jones, cost and schedule overruns can be prevented by a combination of three techniques: 1) improve the accuracy of initial cost estimates, 2) improve software

engineering techniques in order to reduce software costs, and 3) accurately measure software cost and schedule metrics throughout the program. The only one of these methods that can be utilized before the start of a program is the use of software cost and schedule estimating methods. Jones also recommends the use of commercial grade software estimating tools. (Jones, 1994:40, 89, 159)

Unfortunately, most estimating techniques make the erroneous assumption that all will go well with a project, and they often confuse effort with progress, i.e., they disregard the partitionability of the task. For example, a perfectly partitionable task that takes one person nine months can be accomplished by nine people in one month. However, an unpartitionable task, that takes one person nine months (e.g., a woman having a baby) cannot be finished in a shorter time no matter how many people are assigned. This is also true for software. Software development tasks can only be partitioned to a certain level, beyond which diseconomies of scale due to communication and integration problems exceed the economies achieved due to partitioning. Also, adding people to a late project will actually make it later, due to the increased communication and training requirements. (Brooks, 1975:14-26)

Even without considering these realities, the typical uncalibrated model is doing well if it can estimate within 20% of the actual cost 70% of the time when it is used in the area from which it is derived; outside of this area it will do much worse (Boehm, 1981:32). The findings of Ourada's thesis indicate that the average accuracy of several uncalibrated cost models in use within the DoD was only within 50% of actual costs (Ourada, 1991:Chapter 4).

Types of Estimation Methods. Acceptable software cost estimating methodologies fall into five basic categories: algorithmic, bottom-up, top-down, expert judgment, and analogy. These methods are not mutually exclusive and may be combined.

Also, none of these techniques is always better than the others, and the strengths and weaknesses of different methods may be complimentary. All of these methods have their advantages and disadvantages as discussed in the following paragraphs. Unacceptable methods include Parkinson, which equates the cost estimate with the available resources (work expands to fill the available volume) and Price to Win, which bases the cost estimate on what is believed will win the contract or job. (Boehm, 1981:329-330, 340)

The algorithmic method, which is sometimes equated with the top-down method, bases its estimate on system level characteristics and is the basis for most commercial software cost estimating models favored by the DoD (Ferens, 1995). A computerized, empirical, parametric model utilizes equations with input and internal parameters that were derived based on analysis of specific databases containing information on past software development projects (Wellman, 1992:36-38). The default values for the parameters, when combined with user inputs for software size, complexity, developer skill, and other factors, result in a forecast (estimate) for software cost and schedule. The advantages of these models include their ease of use, objectivity, consistency, and their usability early in a program, when little specific data are available and default values may be used (Boehm, 1981:332-333, 341; Wellman, 1992:36-38). However, estimates obtained using the default parameter values embedded in these models are usually inconsistent and inaccurate (Boehm, 1981:320-321, 330-333, 341), and the models tend to underestimate the cost, size, and schedule of future projects (Brooks, 1975:14-16). They only produce estimates that are accurate to within 25% about 50% of the time (Ourada, 1991:Chapter 4) due to the age of the databases upon which the equations are based, the appropriateness of the original databases to the current project (i.e., they are database dependent), and the inherent instability of the models. This inherent instability is caused by the fact that small parameter changes can result in large differences in results. Model estimating errors are often due to their use outside of the original environment from which they were derived

and calibrated (Wellman, 1992:33). In other words, the estimation error often resides within the historical data, rather than within the model calculations (Jones, 1994:155). Therefore, calibration to a specific organization and its environment is a must (Wellman, 1992:33).

The bottom-up method involves defining the project to the lowest levels of the Work Breakdown Structure (WBS) and estimating each WBS element. This can result in very accurate estimates due to the greater knowledge of the amount of work to be done and also to the law of large numbers, which states that the errors tend to average out when the units are summed. This method is not used very often in the DoD since it is quite time consuming, the data are not available early in the program, and the data may be proprietary. This method may also underestimate system level costs, such as integration, configuration management, quality assurance and program management costs, since it estimates the lower WBS elements without looking at the system level big picture. Also, with a large WBS, there is a good possibility that items will be omitted. (Boehm, 1981:338-342)

The top-down method derives the cost estimate from the global properties of the software system and splits it among the different components. This captures the system level functions well, but may miss low level WBS elements. It may not unmask low level technical problems that might have been caught by other approaches, such as the bottom-up method. It is also less stable than the bottom-up method, since estimation errors do not have a chance to balance out. (Boehm, 1981:337-338, 342)

The expert judgment method, often referred to as the Delphi technique, relies on the opinions of a sampling of experts in the type of software being estimated. For small projects that are very similar to past projects and that are in the experts' field of knowledge, this may produce an accurate estimate. However, significant errors can occur

due to bias of the experts, knowledge levels of the experts, and dissimilarity to the projects with which the experts are familiar. (Boehm, 1981:333-335, 342)

The analogy method is based on comparisons between the project under development and specific past projects that are very similar. Problems with this method are similar to some of those experienced by the algorithmic method, i.e., old data for new programs and a lack of an appropriate database. (Boehm, 1981:336, 342).

The advantages and disadvantages of the above methods are summarized in Table 1.

Calibration. Some of the estimating models have achieved improved accuracy via calibration (Ferens and Christensen, 1995:15). This improvement in accuracy can be as high as a factor of five (Thibodeau, 1991:5-29). However, the improved accuracy level may still not be very good, especially if the default accuracy was very poor (Ferens and Christensen, 1995:15).

Calibration involves deriving a new set of parameter values by adapting the model to a specific database containing analogous historical software development efforts. This adjusts the model's estimates to the particular environment of the user. Major sources of error that can be calibrated include "error in production rate assumptions", or errors in how fast work can be done, and "errors in assignment scope", or miscalculating the amount of work that a person can perform (Jones, 1994:156-157).

Calibration is an iterative process that refines a model's generic parameters to new parameters that correspond to specific attributes of a software project, such as application type, programming language, and development contractor. The calibration is repeated until the desired level of precision is achieved, e.g., until the calibration constants are refined to three significant figures. The calibration is performed using the results of similar completed projects, in order to improve estimates for future projects. Ideally, the model

should be calibrated to specific contractors, operating environments, application types, and languages.

Table 1. Strengths and Weaknesses of Software Estimation Methods

Method	Strengths	Weaknesses
Algorithmic	• Objective, repeatable, analyzable, formula • Efficient, good for sensitivity analysis • Objectively calibrated to experience	• Subjective inputs • Assessment of exceptional circumstances • Calibrated to past, not present or future
Bottom-Up	• More detailed basis • More stable • Fosters individual commitment	• May overlook system level costs • Requires more effort (more expensive to perform)
Top-Down	• System level focus • Efficient	• Less detailed basis • Less stable
Expert Judgment	• Assessment of representativeness, interactions, exceptional circumstances	• No better than the participants • Biases, incomplete recall
Analogy	• Based on representative experience	• Representativeness of experience
Parkinson	• Correlates with some experience	• Reinforces poor practices
Price to Win	• Often gets the contract	• Dishonest • Generally produces large overruns

(Boehm, 1981:342)

Space and Missile Systems Center (SMC) Software Database (SWDB)

The SMC SWDB is a PC based data retrieval system that uses the Microsoft FoxPro database system. The SMC SWDB, Version 2.1, contains 2638 records of defense related software projects totaling approximately 50 million source lines of code

(SLOC) ("the most extensive software database available in the government"). However, only 444 records have effort information available. The SMC SWDB was developed under the direction of the SMC/FMC (directorate of cost) with assistance from the Space Systems Costs Analysis Group (SSCAG). Each record contains 276 data fields in four sections: 1) general information, 2) cost, size and schedule information, 3) software characteristics, and 4) software maintenance (support) information. A fifth section, proprietary data, was originally included, but has since been removed. Therefore, unfortunately, the records do not contain a field that identifies the developing contractor, not even an anonymously assigned contractor designator or number. This was done to encourage contractors to anonymously contribute accurate information to the SMC SWDB without the fear that it would later be used against them. While this necessary safeguard encourages contractor honesty and increases the size of the SMC SWDB, it makes calibration and data analysis more problematic. (Stukes, 1996:3-11)

The SMC SWDB contains three modes for accessing records. The browse mode accesses the records one at a time in a sequential manner. The find mode accesses the records one at a time by going directly to a specific record number. The query mode accesses multiple records by searching the entire database for records that match certain parameters. Searchable parameters include the software level (project, CSCI, CSC, unit, other), the operating environment (e.g., military ground), the application (e.g., command and control), the software function (e.g., display), and the language. Ranges can also be specified for effective (normalized) size, total size, development effort, and years of maintenance. (Management Consulting and Research, 1995)

The SMC SWDB was developed in five stages. First, the project was designed based on six software cost models: REVIC, SEER-SEM, PRICE S, SoftCost-OO, SLIM, and SASET. Second, data were collected and verified for consistency and reasonableness. Third, the data were mapped and normalized to account for inflation, economies of scale,

technology, design year, new versus upgrade, and incomplete systems. Fourth, the database was automated. Fifth, the project was documented. This development is an on-going process, with new data fields being created, new records being collected and normalized, and new versions of the automated database and user's manual being produced. The data in the SMC SWDB have been obtained from government sources (SMC programs, European Space Agency (ESA), NASA, and Air Force Materiel Command (AFMC)), major aerospace companies, suppliers, non-aerospace companies, software estimating model developers, and other databases (Aerospace, SSCAG, General Dynamics, ESA, JPL, etc.). The SMC SWDB project used several methods to attempt to ensure the reliability of the data; however, much of the data's reliability depends on the honesty and accuracy of the submitting contractors. The data were requested using a standardized form and data dictionary; data from other databases were analyzed to ensure consistent definitions (for example, all SLOC counts are logical lines, rather than physical lines). Also, the data are screened and evaluated using sanity checks and metrics, and data sources are often re-contacted. (Stukes, 1996:3-10)

SMC desires to have several commercial cost estimating models calibrated to a level of specificity that will enable improved estimating accuracy for future software development efforts. It is hoped that this improved accuracy can be achieved by calibrating these models to specific subsets of the SMC SWDB.

To be useful for calibrating a software cost estimating model, each of the records will need to be reviewed to determine its suitability. Suitability is based on consistency of data, the presence of actual cost (effort) data for the completed project and the presence of sufficient information to satisfy the data input requirements of the model to be calibrated.

Variability within most databases makes small error prediction with a single model difficult (Taub, 1993:190). However, subdividing the data by size and type into

homogeneous data sets reduces this variability and allows the estimate to be made at a reasonable probability level (Taub, 1993:190). Therefore, all suitable records will be divided into stratified data sets, based on application type, in order to obtain new parameters that are based on sets of records that are homogeneous enough to yield accurate estimates.

Previous calibrations to the SMC SWDB have been performed for PRICE S (Galonsky, 1995), SLIM (Kressin, 1995), SEER-SEM (Rathmann, 1995), SASET (Vegas, 1995), and REVIC (Weber, 1995). These studies had mixed results, with little substantial improvement in absolute estimate accuracy. They did, however, show that calibration could improve estimates over the default values in some instances. However, "the few calibrations which met Conte's criteria for a good model were based on very few data points." Table 2 shows the model name and owner, thesis author, and the calibration parameters. (Ferens and Christensen, 1995:15)

Table 2. Software Cost Models Calibrated

Model	Responsible Agency	Thesis Author	Calibration Method(s)
REVIC	AF Cost Analysis Agency	Weber	• Coefficient and Exponent • Coefficient Only • SAS - Coefficient and Exponent with no Productivity Multipliers
PRICE S	PRICE Systems	Galonsky	• Productivity Factor (PROFAC)
SEER-SEM	Galorath Associates	Rathmann	• Effort Adjustment Factor
SLIM	QSM, Inc.	Kressin	• Productivity Index (PI)
SASET	U.S. Navy Cost Center	Vegas	• Software Type Multiplier • Class Multiplier (alternate)

(Ferens and Christensen, 1995:4)

Normalization. (Stukes, 1995:F-1 to F-3) The SMC SWDB query reports also provided normalized effective size and normalized effort information. The size was

normalized to account for the differences between new, modified, and reused lines of code. The effort was normalized to account for differences in the contractor supplied effort data for the development phases and person-hours of effort per person-month.

The size normalization procedure used contractor inputs for new, modified, and reused SLOC to compute an equivalent new SLOC figure. This equivalent new SLOC figure was computed as follows:

1. Multiply the percentage of the reused and modified code that was re-designed by the number of SLOC of reused and modified code and by a 40% weighting factor for the design phase.

2. Multiply the percentage of the reused and modified code that was re-coded by the number of SLOC of reused and modified code and by a 25% weighting factor for the coding phase.

3. Multiply the percentage of the reused and modified code that was re-tested by the number of SLOC of reused and modified code and by a 35% weighting factor for the testing phase.

4. Sum the results of steps 1-3 to get the equivalent new code figure for the reused and modified code.

5. Add the result of step 4 to the number of SLOC of new code to get the total equivalent new code figure.

These steps are consolidated in the following equation:

$$\text{Equivalent New SLOC} = (\text{New SLOC}) + (40\% * \text{Reused SLOC} * \%\text{Re-design}) + (25\% * \text{Reused SLOC} * \%\text{Re-code}) + (35\% * \text{Reused SLOC} * \%\text{Re-test}) \quad (1)$$

The effort normalization procedure used contractor inputs for effort, person-hours per person-month, and development phases included in the reported effort. The reported effort was converted to 152 person-hours per person-month by multiplying the reported effort by the reported number of person-hours per person-month and then dividing by 152. It was also scaled from the reported development phases to a SMC SWDB standard

development consisting of the beginning of the preliminary design phase through the end of CSCI testing. This scaling was done based on the following table:

Table 3. SMC SWDB Software Phase Normalization

Phase	% of Normalized Effort
Software Requirements	5.5
Preliminary Design	11.4
Detailed Design	19.1
Code and Unit Test	29.8
CSC Testing and Integration	35.6
CSCI Testing	4.1
Systems Test and Integration	7.2
OT&E	4.8

(Stukes, 1995:F-2)

SoftCost-R Calibration

One of the commercial models that has not yet been calibrated to the SMC SWDB is SoftCost-R. SoftCost-R is a forerunner of SoftCost-Ada and SoftCost-OO. The SoftCost family of cost estimation models (SoftCost-R, SoftCost-Ada, and SoftCost-OO) are the only models used as the basis for the SMC SWDB data collection sheet that have not yet been calibrated to the SMC SWDB. Therefore, further research needs to be performed in order to calibrate the SoftCost family of models to the SMC SWDB. SoftCost-R is the SoftCost model that was chosen due to the relative ease with which it can be calibrated. SoftCost-R's constants are all contained within a calibration file and can be modified with any text editor. The newer SoftCost-Ada and SoftCost-OO models, however, have most of their constants coded into the executable source code where they cannot be changed.

"SoftCost-R is a parametric software cost estimation package, based on the SoftCost mathematical model developed for NASA in 1981 by Dr. Robert C. Tausworthe of the Jet Propulsion Laboratory." SoftCost-R, Version 8.4, incorporates extensions developed by Donald J. Reifer to improve the SoftCost model. These changes were designed to reflect changes to the state-of-the-practice and yield a complete, mature, general purpose software cost estimation package. The current model contains inputs for 33 cost factors (22 natural logarithm multipliers from Tausworthe, 9 linear multipliers from Reifer, software type, and analyst capability), application domain, and system architecture. The equations in SoftCost-R are public domain, and its parameters are based on calibration to RCI's database of over 1500 software projects. (Reifer Consultants, 1989:R-1 to R-2, R-7)

SoftCost-R contains several submodels, including a separate submodel to perform intermediate COCOMO estimates as a sanity check to the SoftCost-R estimates. The SoftCost-R model itself consists of a sizing submodel, an estimating submodel, a risk submodel, an allocation submodel, and a life cycle submodel. The sizing submodel initially calculates an equivalent software size expressed in thousands of source lines of code (KSLOC), called thousands of source lines of executable code (KSLEC) by the SoftCost-R documentation, based on user, ASSET-R, or Software Sizing Model (SSM) inputs for new and reused lines of code and function points. The estimating submodel then calculates base effort and duration estimates. The risk submodel can be used to tradeoff effort and duration to make cost and schedule resource planning decisions. The allocation submodel takes the effort and duration and allocates them to Work Breakdown Structure (WBS) tasks and labor categories. The life cycle submodel extrapolates the data into the Operations and Support phase and allows tradeoffs to be made between cost and support manning levels. The interaction between these submodels is depicted graphically

in Figure 1. This research effort will only be concerned with the estimating submodel. (Reifer Consultants, 1989:R-4 to R-5)

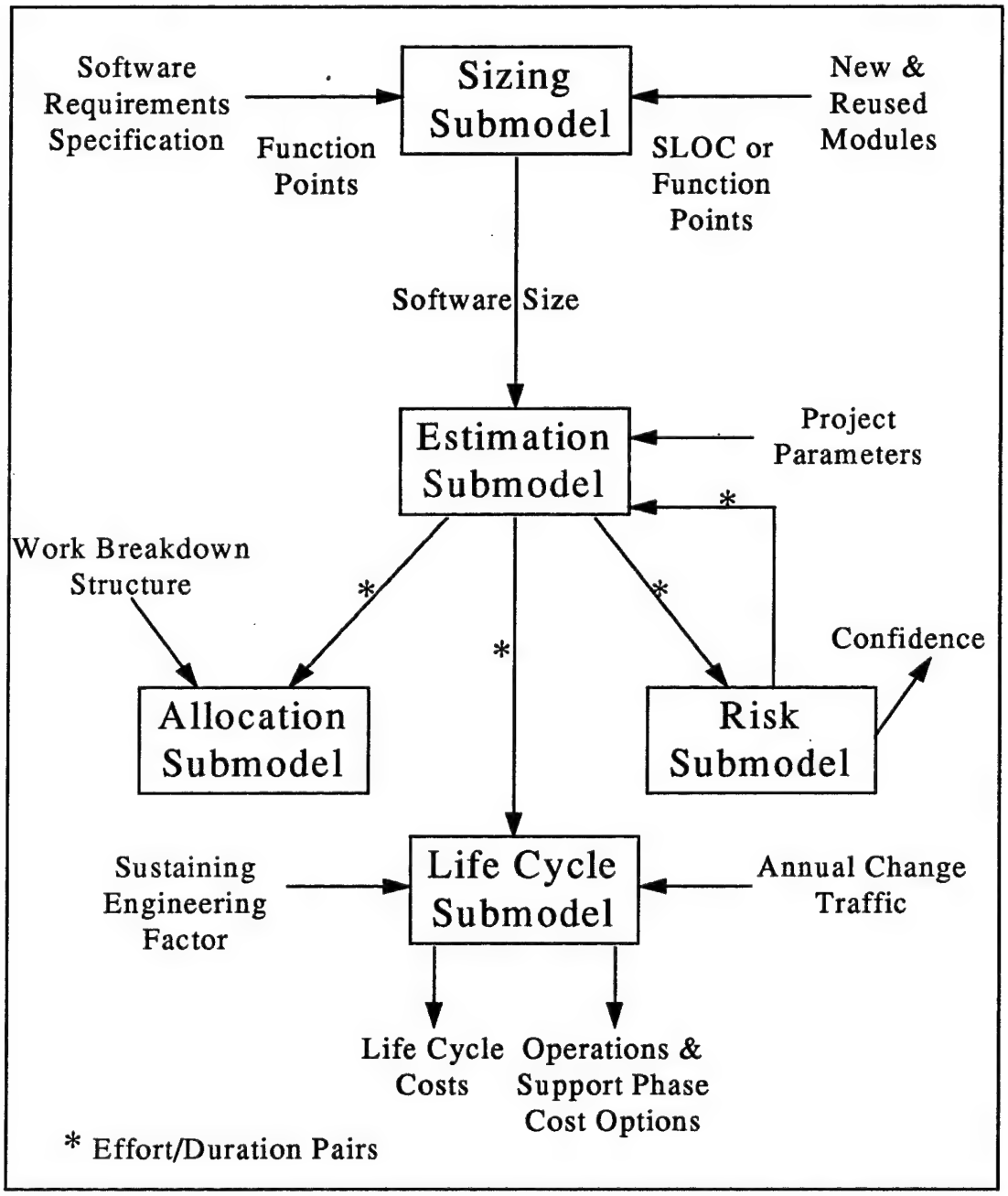

Figure 1. SoftCost-R Submodel Architecture
(Reifer Consultants, 1989:R-6)

The basic mathematical formulation of the estimating submodel is as follows:

$$\text{Effort} = 1.725*\text{GAMMA\_W}*\text{PA}*(\text{Size})^{A} \quad (2)$$

$$\text{Duration} = 1.0*\text{GAMMA\_T}*\text{DFAC}*(\text{Effort})^{\text{DEXP}} \quad (3)$$

where Gamma_W is an application domain specific effort calibration constant, PA is a calculated effort technology constant, Size is in KSLOC, and A is the effort exponent (see Table 4). The constant 1.725 is an effort adjustment that is coded in the SoftCost-R source code, but it is not mentioned in the reference documentation. (Reifer Consultants, 1989:R-10, R-14; Bruscino, 1996)

Since this research is limited to calibration of the software effort, only the effort equation will be detailed further. All variable and constant names given here are from the SoftCost-R documentation. In the cases where these names are different in the SoftCost-R source code, the source code names are included in parenthesis the first time they appear. Gamma_W is obtained from Table 5. The PA parameter includes the cost drivers and a productivity calibration factor (which contains the KSLECPM and AWFFAC parameters to be calibrated by this research). The equation for PA is as follows:

$$\text{PA} = P_0*[A_1*(W_N)^2*\exp(A_2)]^{\text{AWF}} \quad (4)$$

where P_0 is a productivity factor, A_1 (P_η) is the product of the nine Reifer linear multipliers, W_N (Effort_11) is an effort normalization factor, A_2 (index) is the sum of the twenty-two Tausworthe natural logarithm multipliers, and AWF is the work effort tradeoff

factor. The factors of A_1 are given in Table 6. All factors of A_1 default to 1.0. The equation for P_0 is as follows:

$$P_0 = [HADJ*EADJ*FULLUP*FADJ*(1+LADJ+TADJ+CADJ)]^{1/2}/KSLECPM \quad (5)$$

where the numerator is approximately equal to 1.771 and is based on seven other constants in the SoftCost-R calibration file (softcost.cal) that are not affected by this calibration method. The seven constants and their default values are given in Table 4. KSLECPM is the productivity ratio factor that will be calibrated by this research effort. The default value for KSLECPM is 0.367 (see Table 4). W_N incorporates the analyst capability factor and adjusts the results for the fact that two of the parameters included in A_2 do not have nominal values that result in unity. The equation for W_N is as follows:

$$W_N = W_C*ANL\_CAP \quad (6)$$

$$W_C = 1.9045/[HADJ*EADJ*FULLUP*FADJ*(1+LADJ+TADJ+CADJ)]^{1/2} \quad (7)$$

where W_C is an effort normalization constant approximately equal to 1.076, and ANL_CAP is the user input value for analyst capability. W_C is based on seven constants that are given in Table 4. The factors of A_2 are given in Table 6. All factors of A_2, except for requirements volatility and adaptation requirements, default to 0.0, which has the effect of multiplying by 1.0, since exp(0.0) = 1.0. Requirements Volatility defaults to 0.131 and Adaptation Requirements defaults to -0.405. The equation for AWF is as follows:

$$AWF = AWFFAC/Index0 \quad (8)$$

$$Index0 = 23.837+\ln(HADJ/EADJ)-\ln(FULLUP)+\ln(FADJ)+\ln(1+LADJ+TADJ+CADJ) \quad (9)$$

where AWFFAC is the productivity adjustment constant that will be calibrated and Index0 is a calculated ln-ratio sum that is approximately equal to 27.259 and is also based on seven constants that are given in Table 4. The default value for AWFFAC is 2.16 (see Table 4). The constant 23.837 is used by the SoftCost-R source code, but its numerical value is not given in the reference documentation. (Reifer Consultants, 1989:F-1 to F-27; Bruscino, 1996)

Table 4. Parameter Default Values

Constant	Default Value
A	1.0
AWFFAC	2.16
KSLECPM	0.367
HADJ	1.2
EADJ	0.8
FULLUP	0.4
FADJ	2.3
LADJ	0.82
TADJ	0.72
CADJ	1.01

(Reifer Consultants, 1989:E-5)

Table 5. Values of Gamma_W

Application Domain	Gamma_W
Automation	1.0
Avionics (includes space)	1.3
Command & Control	1.0
Data Processing	1.0
Environment/Tool	0.8
Scientific	1.0
Simulation	0.9
Telecommunication	1.0
Test	1.0
Other	1.0

(Reifer Consultants, 1989:F-27)

Table 6. Factors of A_1 and A_2

A_1 Component Parameters	A_2 Component Parameters
System Architecture	User Involvement
Staff Resource Availability	Organizational Interface Complexity
Degree of Standardization	Computer Resource Availability
Scope of Support	Security Requirements
Use of Modern Software Methods	Concurrent Hardware Development
Use of Software Tools/Environments	Percentage Code Delivered
Software Tool/Environment Stability	Life Cycle Coverage
Degree of Optimization	Use of Peer Reviews
Database Size	Geographical Co-Location
	Program Complexity
	Database Complexity
	Requirements Complexity
	Requirements Volatility
	Degree of Real-Time
	Adaptation Requirements
	Programmer Capability
	Applications Experience
	Language Experience
	Environment Experience
	Methodology Experience
	Customer Experience
	Team Capability

(Reifer Consultants, 1989:F-10, F-11, F-14)

SoftCost-R uses SLOC internally; all function points are converted to SLOC before the model begins its calculations (Reifer Consultants, 1989:U-45). The minimum and maximum software sizes for projects that can be estimated by SoftCost-R are 5 KSLOC and 3,000 KSLOC, respectively (Reifer Consultants, 1989:U-46). It is claimed that both SoftCost-R and its COCOMO-R submodel estimate within 50% of actual cost and duration 60% of the time (Reifer Consultants, 1989:U-77). The 50% confidence (most likely) estimates reported by SoftCost-R's estimating submodel are actually based on approximately 70% confidence in cost and 70% confidence in schedule, for a 50% confidence that **both** cost and schedule will be achieved (Reifer Consultants, 1989:U-82). The confidence level is defined as the probability that the project will be completed for that amount of effort or less. For example, if a project has a 70% confidence level estimate of 500 person-months, then it will have a 70% probability of costing $\leq$ 500 person-months.

Theory of SoftCost-R Calibration

(Resource Calculations, Inc., undated:1-5)

The KSLECPM and AWFFAC parameters in the SoftCost-R software cost estimating model should be calibrated to specific stratified data sets obtained from the SMC SWDB (Resource Calculations Inc., undated:1). By calibrating the KSLECPM and AWFFAC parameters for specific operating environments and application types, this research aims to improve Conte's goodness of fit for estimates using the calibrated parameters relative to estimates using the default parameters (Conte, Dunsmore, and Shen, 1986:276). New values for these two parameters are obtained by calibration to a sample set of completed projects. During calibration, 11 values from 0.5 to 3.00 in 0.25 increments are selected for AWFFAC. The initial value for KSLECPM is 0.2. The initial values of these constants are based on work done by Interstate Electronics Corporation

(IEC) for RCI. Note that the abbreviations used in the equations were changed to avoid confusion with other abbreviations in this paper. (Collins, 1996)

The general method to be used for each AWFFAC value is as follows:

1. Calculate predicted effort values for each data point based on the AWFFAC value and the initial KSLECPM value.
2. Based on all of the data points in the data set, calculate a new KSLECPM value. This new KSLECPM value should result in more accurate estimates than the initial KSLECPM value.
3. For this AWFFAC and new KSLECPM pair, calculate the sum of the squared differences for the data set.

Repeat the above three steps for each of the eleven AWFFAC values ranging from 0.5 to 3.00. Select the AWFFAC and new KSLECPM pair with the lowest sum of the squared differences as the new calibration values for this data set. This AWFFAC and new KSLECPM pair will be the least squares (best fit) solution for the AWFFAC values used, although it will not be the best least squares solution possible for all AWFFAC and KSLECPM pair possibilities. The range of AWFFAC values from 0.5 to 3.00 was selected based on IEC's empirical data because most of the least squares solutions for their data sets fell in this region. If this method results in a value on the boundary, AWFFAC = 0.5 or AWFFAC = 3.00, the range may need to be extended to find a better least squares solution. (Collins, 1996)

The specifics of this method will now be explained. Step 1 is accomplished by modifying the SoftCost-R calibration file (softcost.cal) to include the initial AWFFAC and KSLECPM pair. The SoftCost-R model is then run to calculate an initial predicted effort (IPE) value for each of the projects in the data set. Step 2 is performed using these IPE estimates and the actual effort (AE) values for each project. A new KSLECPM value is calculated for the data set using the following formula:

$$\text{New KSLECPM} = \Sigma[(0.2*\text{IPE})^2]/\Sigma(0.2*\text{IPE}*\text{AE}) \quad (10)$$

The softcost.cal file is modified to the new KSLECPM value, and the model is run again for each project in the data set. These new predicted effort (NPE) estimates are then compared with the actual effort values to obtain a sum of the squared differences for the AWFFAC and new KSLECPM pair. This sum of the squared differences is obtained using the following formula:

$$\Sigma[(\text{AE}-\text{NPE})^2] \quad (11)$$

The AWFFAC and KSLECPM pair with the smallest sum of the squared differences represents the calibrated values. (Resource Calculations Inc., undated:4)

If it is desirable to fine tune the parameters, use the AWFFAC and new KSLECPM pair for the two smallest sum of the squared differences. Determine the AWFFAC that is halfway between these two AWFFAC values and calculate its new KSLECPM value and sum of the squared differences. Continue this method to halve the difference between the AWFFAC values for the two smallest sum of the squared differences from the previous iteration until the desired AWFFAC and KSLECPM precision is achieved Usually three times is sufficient. (Resource Calculations Inc., undated:4)

Equation (12) is a modification of the original IEC formula that was given in the RCI paper. The original IEC formula,

$$\text{New KSLECPM} = \Sigma(\text{AE}*\text{IPE}/0.2)/\Sigma[(\text{IPE}/0.2)^2] \quad (12)$$

was in error. When the original formula is used, the new KSLECPM value makes the estimate worse, not better. If the model is underestimating the effort, the original formula increases the KSLECPM value which lowers the effort estimate even further (since KSLECPM is a divisor in the effort equation). Similarly, if the model is overestimating the effort, the original formula decreases the KSLECPM value and raises the effort even further.

Summary

Throughout the literature it was stressed that the importance and expense of software in DoD projects makes accurate cost estimating a must in the current environment of tight budgets and management scrutinization. In order to perform these estimates early in a program, the only objective tools that have shown consistent results are the algorithmic cost estimating models. On the other hand, studies and experience have shown that even these models have not exhibited a desired level of accuracy. However, some of these models have shown improved results when calibrated to databases that are specific to the projects to be estimated.

One algorithmic cost model that could be calibrated to the SMC SWDB is SoftCost-R. The accuracy of the new calibrated inputs should be verified through comparisons between the calibrated and default estimates obtained from the model and the actual cost data contained in the SMC SWDB. As other projects are added to the SMC SWDB in the future, the new parameters should be verified by running the model and comparing to the actual cost data. Also, as software productivity changes, the model will need to be re-calibrated periodically in the future.

In conclusion, uncalibrated software models have not proven to be accurate for a wide range of operating environments, applications, and languages. However, these models can be calibrated to specific databases in an attempt to produce better estimates.

This calibration has been shown to improve accuracy relative to the uncalibrated models, especially when the calibration set is fairly homogeneous. Therefore, SoftCost-R should be calibrated to subsets of the SMC SWDB that are as homogeneous as possible while still maintaining statistically significant sample sizes.

III. Methodology

Overview

The methodology used here to calibrate SoftCost-R was based on the IEC method described in Chapter II of this thesis. The model was calibrated using the SMC SWDB of defense industry software development projects. Records containing data suitable for calibrating SoftCost-R were extracted and stratified into multiple data sets, based on their operating environment and application type. These data sets were further divided into calibration and validation data subsets. The goal was to use the calibration data sets to calculate new parameters to be applied to future DoD software development efforts. These new parameters were validated using the validation data sets. Statistical analysis performed for this validation included MRE, MMRE, RMS, RRMS, Pred (0.25), and Wilcoxon signed-rank tests.

Procedures and Data Analysis

The model was calibrated by deriving new productivity ratio factors (KSLECPM) and productivity adjustment constants (AWFFAC) that will be applicable to defense software industry efforts in specific application areas. The KSLECPM is a direct linear multiplier of the effort required to complete the project; it is a coefficient of PA (see Equations (2), (4), and (5)). AWFFAC is a multiplier in the exponent used to calculate PA (see Equations (2), (4), and (8)), and it has a logarithmic effect on the effort. The operating environments and application areas calibrated were determined by the presence of suitable records in the SMC SWDB. A preliminary list included military ground command and control, military ground signal processing, unmanned space, ground in support of space, military mobile, missile, and military specification avionics operating environments and application types.

The first step was to research the SoftCost-R model to obtain a thorough understanding of its functionality, calculations, and parameter sensitivities. This step included an analysis to verify that the productivity ratio factor and productivity adjustment constant were suitable for calibration.

Next, the SMC SWDB was searched for records of software projects that contained enough information to run the SoftCost-R model. The required information included CSCI level reporting, operating environment and application type in one of the desired categories, effort greater than zero, size between 5 and 3,000 KSLOC, country (U.S. or Europe), and nominal or higher data confidence. This search of the SMC SWDB was performed using the query feature of the database. Each query was saved in a database format (*.dbf) that could be read by Microsoft Excel, Version 5.0. The unique information for the seven queries that were performed is listed in Table 7. The following information was common to all seven of the queries that were performed:

1. Software Level = CSCI
2. Software Functions = All
3. Programming Language = All
4. Effective Size Range = 5,000 to 3,000,000 (not including records with this field empty)
5. Total Size Range = 0 to 9999999999 (including records with this field empty)
6. Effort Range = 0 to 9999999999 (not including records with this field empty)
7. Years of Maintenance = 0 to 9999999999 (including records with this field empty)

Table 7. SMC SWDB Queries

Query Title	Operating Environment	Applications
Mil Ground - C&C	Mil Ground	Command and Control
Mil Ground - SP	Mil Ground	Signal Processing
Unmanned Space	Unmanned Space	All
Ground in Support of Space	Ground in Support of Space	All
Military Mobile	Military Mobile	All
Missile	Missile	All
Mil-Spec Avionics	Mil-Spec Avionics	All

Each query was divided into two data sets, one for U.S. projects and one for European projects. Each data set was then randomly divided into two subsets using the random number generator in Microsoft Excel, Version 5.0. One subset was used for calibration of the model, while the other subset was used for validation of the new calibration constants. If there were seven or fewer total data points, calibration and validation were not performed, since the small sample size would call into doubt the statistical validity of the results. For eight or more total data points, one-half of the data points (rounded up) were included in the calibration data set, and one-half of the data points (rounded down) were included in the validation data set.

Rather than following the IEC calibration method, which requires running the SoftCost-R model a minimum of 22 times for each calibration data point and once for each validation data point, the SoftCost-R effort calculations were emulated using Microsoft Excel, Version 5.0. The Microsoft Excel spreadsheet uses the database records, the SoftCost-R equations, the softcost.cal calibration file values, and a set of default factor values derived from the SoftCost-R quick run option and based on the software type. The SMC SWDB fields that the spreadsheet uses to calculate each SoftCost-R input factor are listed in Appendix B. The spreadsheet simultaneously calculates the initial predicted effort (IPE) for the entire calibration data set. The solver

function then determines the optimum AWFFAC and KSLECPM pair to provide a "best fit" solution to the calibration data set. It should be noted that the assumptions of ordinary least squares regression, i.e., normality, common variance, and independence, were not tested. The spreadsheet then calculates the predicted effort for the records in the validation data set and calculates the statistical parameters used to validate the calibration. The ability of the spreadsheet to accurately emulate the SoftCost-R model was verified by both a manual code walk through and the use of default data records that were run through both the spreadsheet and the SoftCost-R model.

For each data set, the SMC SWDB query file was copied into the spreadsheet, and the following steps were performed:

1. Each record in the data set was marked with either a "C" for a calibration record or a "V" for a validation record.

2. The software type was entered into the spreadsheet (see Table 8). The SoftCost-R quick run factor values were used as the default values for data fields that were empty. The only changes to the quick run factor values between the model and the spreadsheet were to change the Degree of Standardization default from 1.0 (commercial standards) to 1.21 (tailored military standards) for the TELECOMM software type and to change the Life Cycle Coverage default from 0.25 (System Requirements Review through the end of System Test) to -0.375 (Software Specification Review through the end of Software Test) for all of the software types. The first change was made due to the fact that most of the Military Ground - Signal Processing (TELECOMM) data records were developed for the military at a time when military standards were mandatory. The second change was made due to the fact that the SMC SWDB effective size parameter was used for each data record, and this effective size parameter is normalized to the phases from preliminary design through CSCI test (Stukes, 1995:F-2).

3. The degree of standardization data field was updated manually for each data record that had information in this field (non-empty field).

4. The solver data analysis function was run on the calibration records to minimize the sum of the squared differences (see Equation (11)) by having solver adjust the AWFFAC and KSLECPM parameter values. The initial AWFFAC value was 2.16 (the model default), and it was allowed to vary from 0.0 to 10.00. The range was set wider than the IEC calibration method of 0.5

to 3.00 in order to encompass more of the true minimum least squares solutions, yet it was limited to a maximum of 10.00 to prevent unnecessarily long solver solution calculations. The initial KSLECPM value was 0.367 (the model default), and it was restricted to positive real numbers. The resulting AWFFAC and KSLECPM values were calculated to a precision of 0.001.

Table 8. Software Type

Query Type	SoftCost-R Quick Run Type	SoftCost-R Quick Run Name
Mil Ground - C&C	Command and Control	CMD_CTRL
Mil Ground - SP	Telecommunications	TELECOMM
Unmanned Space	Unmanned Space	UNMANSPC
Ground in Support of Space	Mission Critical Ground	MCRITGND
Military Mobile	Mission Critical Ground	MCRITGND
Missile	Unmanned Space	UNMANSPC
Mil-Spec Avionics	Mission Critical Air	MCRITAIR

To validate the hypothesis that calibration improves the estimating accuracy of SoftCost-R, the calibrated and default effort estimates for the validation data set were compared to the actual effort values, using various statistical methods, to determine the amount of improvement achieved by the calibration. The validation statistics were obtained from the Data Summary worksheet of the spreadsheet. The results were summarized for each application type and reported in Chapter IV of this thesis. The statistical analysis in the spreadsheet included the following equations which are also summarized in Table 9:

$$\text{Magnitude of Relative Error (MRE)} = |(\text{Estimate-Actual})/\text{Actual}| \quad (13)$$

$$\text{Mean Magnitude of Relative Error (MMRE)} = (\Sigma\text{MRE})/n \quad (14)$$

$$\text{Root Mean Square} = \{(1/n)*\Sigma[(\text{Estimate-Actual})^2]\}^{1/2} \quad (15)$$

$$\text{Relative Root Mean Square (RRMS)} = \text{RMS}/[(\Sigma\text{Actual})/n] \quad (16)$$

$$\text{Prediction Level (Pred (0.25))} = (k/n)*100\% \quad (17)$$

where k is the number of data points with MRE $\leq$ 25%. A good fit, or "acceptable model performance", occurs when a model consistently produces estimates that satisfy all of the following criteria (Conte, Dunsmore, and Shen, 1986:276):

$$\text{MMRE} \leq 25\% \quad (18)$$

$$\text{RRMS} \leq 25\% \quad (19)$$

$$\text{Pred (0.25)} \geq 75\% \quad (20)$$

where at least 75% of the projects are predicted within 25% of their actual results. However, heterogeneous data sets with widely differing complexity and size make it difficult to satisfy both Equations (18) and (19). Therefore, satisfaction of Equations (18) and (20) may be more reasonable (Conte, Dunsmore, and Shen, 1986:276). The distributions of the calibrated and default estimates were also compared to the distributions of the actual results, using a Wilcoxon non-parametric statistic, in order to check for bias (see Appendix E).

The MRE indicates the degree of estimating error in an individual estimate. It is calculated by using Equation (13). As MRE decreases, the estimate is more accurate (i.e., it has less error). The MMRE indicates the average degree of estimating error in a data

set. It is calculated using Equation (14). The lower the MMRE value, the better the model represents the data as a whole (lowest average error). The RMS reflects the model's ability to accurately forecast the individual actual effort, and it is calculated by using Equation (15). The RRMS represents the model's ability to accurately forecast the average actual effort; it is calculated using Equation (16). Pred (0.25) is the percentage method. It is used to validate the predictive ability of a model. It gives the percentage of estimates that are within 25% of the actual results. It is calculated using Equation (17) and is especially useful when a few extreme outliers throw off the other statistics, since each record is weighted the same regardless of how accurately it was estimated. Note that the MRE, MMRE, RRMS, and Pred (0.25) values are useful for comparisons between data sets, since they are relative measurements, while the RMS statistic is only useful within the same data set, since it is an absolute measure.

Table 9. Statistics Summary

Statistic	What it Shows	Goodness of Fit Criteria
MRE	The degree of estimating error in an individual estimate	Lower is better
MMRE	The average degree of estimating error in a data set	MMRE $\leq$ 25%
RMS	The model's ability to accurately forecast the individual actual effort	Lower is better
RRMS	The model's ability to accurately forecast the average actual effort	RRMS $\leq$ 25%
Pred (0.25)	The percentage of estimates that are within 25% of the actual results	Pred (0.25) $\geq$ 75%

(Conte, Dunsmore, and Shen, 1986:276)

Summary

To summarize, appropriate records from the SMC SWDB were divided into data sets stratified by operating environment and application type. Each data set was randomly divided into two subsets, one for calibration and one for validation. The SoftCost-R model was emulated in a spreadsheet and new AWFFAC and KSLECPM values were determined from the calibration data sets. The calibrated and default parameters were then used to obtain calibrated effort estimates and default effort estimates for the validation data sets. To validate the calibration, the calibrated and default effort estimates were compared to the actual effort data using several statistical methods to determine what improvement in the model's estimates resulted from the calibration.

IV. Findings

Overview

This chapter presents the analysis and results of the research. It details the stratification of the SMC SWDB data records, the results of the calibration for each data set, and the results of the validation for each data set. It also discusses the statistical analysis performed on the validation data sets.

SMC SWDB Stratification Results

The goal was to select acceptable software efforts from the SMC SWDB and sort them into homogeneous data sets by environment and application type. This selection and sorting procedure was performed using the SMC SWDB's query feature. Each query was specific to a certain environment and application type and only selected records of software efforts at the CSCI level. The queries did not include records with no effort, and they eliminated records with less than nominal data confidence, less than 5 KSLOC, or greater than 3000 KSLOC. The data confidence level used is a field reported by the SMC SWDB report generator to estimate the confidence in the normalized size and effort data. This confidence level is based on the amount and consistency of the new software size, pre-existing software size, % re-design, % re-code, % re-test, and software development phases data that is provided. The confidence level is an indicator of how likely the SMC SWDB normalized data accurately represents the true normalized size and effort. Higher confidence levels represent normalized estimates based on complete and consistent data; lower confidence levels represent normalization estimates based on incomplete or inconsistent data. Of the data records that met all other criteria, only three were excluded for less than nominal data confidence level. (Stukes, 1995:F-1 to F-3)

In order to obtain the maximum possible amount of homogeneity in the data sets, it was decided to create separate data sets for U.S. and European records. This was done due to the fact that the European efforts may have used different development standards, techniques, and processes. Only two queries contained at least eight European efforts, the minimum number required for calibration and validation. These two categories were Ground in Support of Space and Unmanned Space. For these two categories, the query was split into two separate data sets, one for European developments and one for U.S. developments. For the other categories, the European records were eliminated. The missiles category also did not have enough records to perform calibration and validation, and it was eliminated from further study. These procedures resulted in six data sets of U.S. records and two data sets of European records. These eight homogeneous data sets were then split into calibration and validation sets for each environment and application type. A summary of these data sets is given in Table 10, and a listing of the records is given in the calibration and validation sections of this chapter.

SoftCost-R Calibration Results

Calibrating the SoftCost-R model to the SMC SWDB involved changing the AWFFAC and KSLECPM parameters until a least squares (best fit) solution was found for each calibration data set. In the uncalibrated SoftCost-R model, the default value for AWFFAC is 2.16 and the default value for KSLECPM is 0.367. The results of this calibration for each data set are given in the following sections. The complete listing of all of the records in each calibration data set, including size, default effort, calibrated effort, and MRE, is located in Appendix C. The MRE range and the Pred (0.25) results give a rough idea of how well the model fit the calibration data set. The smaller the calibrated MRE range and the larger the calibrated Pred (0.25), the better the fit the calibration was able to achieve between the SoftCost-R model and the calibration data set. Also, the

greater the improvement between the default MRE range and Pred (0.25) and the calibrated MRE range and Pred (0.25), the better the expected improvement due to the calibration.

Table 10. SMC SWDB Query Results

Query Title	Operating Environment	Applications	Number of Acceptable Data Points	Number Used For Calibration	Number Used For Validation
Mil Ground - Command & Control	Mil Ground	Command & Control	12 U.S.	6	6
Mil Ground - Signal Processing	Mil Ground	Signal Processing	19 U.S.	10	9
			1 European	0	0
Unmanned Space	Unmanned Space	All	11 U.S.	6	5
			15 European	8	7
Ground in Support of Space	Ground in Support of Space	All	30 U.S.	15	15
			50 European	25	25
Military Mobile	Military Mobile	All	10 U.S.	5	5
Missile	Missile	All	4 U.S.	0	0
Mil-Spec Avionics	Mil-Spec Avionics	All	10 U.S.	5	5

Military Ground - Command and Control. The military ground command and control data set consisted of 12 useable data records. Therefore, six of these records were randomly selected to calibrate SoftCost-R to this environment and application type. The records and the results of this calibration are given in Appendix C. The calibrated AWFFAC is 0.676, and the calibrated KSLECPM is 0.730. The default effort MRE ranged from 82.00% to 280.83%, and the calibrated MRE ranged from 1.01% to 67.13%. The default Pred (0.25) was 0.00%, and the calibrated Pred (0.25) was 83.33%. This calibration was able to improve the estimate of every record in the calibration data set. This fact, along with the signs of the differences, indicates that this data set was quite

homogeneous, and the calibration was primarily adjusting for the overestimation tendency of SoftCost-R with the default parameters.

Military Ground - Signal Processing. The military ground signal processing data set consisted of 20 useable data records. The single European effort (Record 2592) was discarded, leaving 19 U.S. records. Therefore, ten of these records were randomly selected to calibrate SoftCost-R to this environment and application type. The records and the results of this calibration are given in Appendix C. The calibrated AWFFAC is 1.712, and the calibrated KSLECPM is 0.685. The default effort MRE ranged from 22.91% to 9884.50%, and the calibrated MRE ranged from 11.53% to 5282.81%. The default Pred (0.25) was 10.00%, and the calibrated Pred (0.25) was 30.00%. This calibration only improved the estimates for six of the ten records in the calibration data set. This fact, along with the signs of the differences, indicates that records were not consistently being over or under estimated by the default model and the accuracy gained due to default overestimates was mostly lost due to default underestimates.

Unmanned Space. The unmanned space data set consisted of 26 useable data records. Eleven of these records were U.S. developments, and six of these records were randomly selected to calibrate SoftCost-R to this environment and application type. Fifteen of these records were European developments, and eight of these records were randomly selected to calibrate SoftCost-R to this environment and application type. The records and the results of these two calibrations are given in Appendix C. The U.S. calibrated AWFFAC is 10.000, and the calibrated KSLECPM is 0.881. The U.S. default effort MRE ranged from 14.90% to 118.27%, and the calibrated MRE ranged from 30.37% to 268.36%. The U.S. default Pred (0.25) was 60.00%, and the calibrated Pred (0.25) was 0.00%. Note that the U.S. calibration only improved the estimates for two of the six records in its calibration data set. This fact, along with the signs of the differences,

indicates that although the records were consistently being underestimated by the default model, the accuracy gained due to default underestimates was mostly lost to an overshoot effect that changed default underestimates to even larger calibrated overestimates for three of the records. Also note that the U.S. calibration was stopped at the upper boundary of AWFFAC = 10. Further calibration at this boundary was generating only minimal improvements in the least squares solution accuracy. The European calibrated AWFFAC is 1.965, and the calibrated KSLECPM is 0.471. The European default effort MRE ranged from 15.66% to 563.78%, and the calibrated MRE ranged from 7.27% to 400.73%. The European default Pred (0.25) was 37.5%, and the calibrated Pred (0.25) was 50.00%. The European calibration improved the estimates for six of the eight records in its calibration data set. This fact, along with the signs of the differences, indicates that this data set was quite homogeneous, and the calibration was primarily adjusting for the overestimation tendency of SoftCost-R with the default parameters.

Ground in Support of Space. The ground in support of space data set consisted of 80 useable data records. Thirty of these records were U.S. developments, and fifteen of these records were randomly selected to calibrate SoftCost-R to this environment and application type. Fifty of these records were European developments, and twenty-five of these records were randomly selected to calibrate SoftCost-R to this environment and application type. The records and the results of these two calibrations are given in Appendix C. The U.S. calibrated AWFFAC is 2.901, and the calibrated KSLECPM is 0.553. The U.S. default effort MRE ranged from 6.94% to 4280.88%, and the calibrated MRE ranged from 1.48% to 3041.56%. The U.S. default Pred (0.25) was 26.67%, and the calibrated Pred (0.25) was 60.00%. Note that the U.S. calibration only improved the estimates for 12 of the 15 records in its calibration data set. This fact, along with the signs of the differences, indicates that this data set was quite homogeneous, and the calibration

was primarily adjusting for the overestimation tendency of SoftCost-R with the default parameters. The European calibrated AWFFAC is 1.597, and the calibrated KSLECPM is 1.095. The European default effort MRE ranged from 4.83% to 808.09%, and the calibrated MRE ranged from 8.19% to 186.52%. The European default Pred (0.25) was 12.00%, and the calibrated Pred (0.25) was 0.00%. The European calibration improved the estimates for 18 of the 25 records in its calibration data set. This fact, along with the signs of the differences, indicates that this data set was fairly homogeneous, and the calibration was primarily adjusting for the overestimation tendency of SoftCost-R with the default parameters.

Military Mobile. The military mobile data set consisted of ten useable data records. Therefore, five of these records were randomly selected to calibrate SoftCost-R to this environment and application type. The records and the results of this calibration are given in Appendix C. The calibrated AWFFAC is 7.842, and the calibrated KSLECPM is 1.083. The default effort MRE ranged from 37.62% to 115.52%, and the calibrated MRE ranged from 0.35% to 32.16%. The default Pred (0.25) was 0.00%, and the calibrated Pred (0.25) was 80.00%. This calibration was able to improve the estimate of every record in the calibration data set. This fact, along with the signs of the differences, indicates that this data set was quite homogeneous, and the calibration was primarily adjusting the estimates closer to the actual values, and that the default model was not over or underestimating.

Missile. Although there were enough suitable data records in this category to perform calibration, there would have been no records for validation. It was decided not to pursue calibration without enough data for validation. Therefore, calibration and validation were not performed. The unused records were Record 15, Record 16, Record 27, and Record 36.

Military Specification Avionics. The military specification avionics data set consisted of ten useable data records. Therefore, five of these records were randomly selected to calibrate SoftCost-R to this environment and application type. The records and the results of this calibration are given in Appendix C. The calibrated AWFFAC is 0.000, and the calibrated KSLECPM is 0.349. The default effort MRE ranged from 19.72% to 146.03%, and the calibrated MRE ranged from 18.51% to 53.97%. The default Pred (0.25) was 20.00%, and the calibrated Pred (0.25) was 20.00%. This calibration was able to improve the estimate of four out of five records in the calibration data set. This fact, along with the signs of the differences, indicates that this data set was fairly homogeneous, and the calibration was primarily adjusting for the overestimation tendency of SoftCost-R with the default parameters. Note that the calibrated AWFFAC value of 0.000, which appears in an exponent of the SoftCost-R equations, effectively eliminates the effect of all of the input factors other than effective size. Therefore, the effort estimate becomes:

$$\text{Effort} = 1.725*\text{GAMMA\_W}*P_0*(\text{Size})^{A} \qquad (21)$$

with Gamma_W being a constant including the avionics software type and P_0 being a constant that includes the calibrated KSLECPM. Since size is the only effective input variable, accurate size estimates become even more essential in order to obtain accurate effort estimates.

Calibration Summary. A summary of the calibration results is given in Table 11. One item of special interest is the fact that only two of the eight calibration data sets achieved Pred (0.25) $\geq$ 75%. Since this predictive level was not consistently achieved,

even when using the calibration data sets from which the calibration parameters were derived, it is not likely that it will be achieved for the validation data sets.

Table 11. Calibration Results Summary

Environment and Application Type	Calibrated AWFFAC	Calibrated KSLECPM	Default MRE Range (%)	Calibrated MRE Range (%)	Default Pred (0.25) (%)	Calibrated Pred (0.25) (%)
Mil Ground - C&C	0.676	0.730	82.00-280.83	1.01-67.13	0.00	83.33
Mil Ground - SP	1.712	0.685	22.91-9884.50	11.53-5282.81	10.00	30.00
Unmanned Space - U.S.	10.000	0.881	14.90-118.27	30.37-268.36	60.00	0.00
Unmanned Space - European	1.965	0.471	15.66-563.78	7.27-400.73	37.50	50.00
Ground in Support of Space - U.S.	2.901	0.553	6.94-4280.88	1.48-3041.56	26.67	60.00
Ground in Support of Space - European	1.597	1.095	4.83-808.09	8.19-188.52	12.00	0.00
Military Mobile	7.842	1.083	37.62-115.32	0.35-32.16	0.00	80.00
Missile	N/A	N/A	N/A	N/A	N/A	N/A
Mil-Spec Avionics	0.000	0.349	19.72-146.03	18.51-53.97	20.00	20.00

SoftCost-R Validation Results

Successful calibration of the SoftCost-R model depended on calculating new AWFFAC and KSLECPM values in order to provide a best fit solution to each of the data sets. These new AWFFAC and KSLECPM parameters were then used to calculate effort estimates for a validation set of records for each environment and application type. These validation set effort estimates were evaluated using the statistical techniques described in Chapter III in order to validate the new calibrated AWFFAC and KSLECPM values. This validation was performed by analyzing the statistical results to see if they support the hypothesis that calibration of SoftCost-R would improve the accuracy of its effort

estimates. The complete listing of all of the records in each validation data set is given in Appendix D.

The range of the MRE statistic, which measures estimation accuracy for an individual data record, was used to determine if the estimating accuracy was becoming more centralized or more dispersed. The MMRE, RMS, RRMS, and Pred (0.25) statistics were used to determine how accurately the model estimated the entire data set. Note that the MRE, MMRE, RRMS, and Pred (0.25) values are useful for performing comparisons between data sets, since they are relative measurements, while the RMS statistic is only useful for analysis within the same data set, since it is an absolute measure. Finally, the Wilcoxon signed-rank test was conducted on the validation data sets for both the default and calibrated signed relative errors to determine if the default and calibrated models are biased toward estimating too high or too low for each particular data set.

Military Ground - Command and Control. The military ground command and control validation data set consisted of six data records. The records and the results of this validation are given in Appendix D. The default effort MRE ranged from 22.48% to 443.09%, and the calibrated MRE ranged from 1.60% to 128.39%. The default effort MMRE was 189.52%, and the calibrated MMRE was 51.86%. The default effort RRMS was 343.33%, and the calibrated RRMS was 87.04%. Since the calibrated MRE range was less varied and closer to the ideal of 0%, the calibrated MMRE was much closer to the ideal of $\leq 25\%$, and the calibrated RRMS was much closer to the ideal of $\leq 25\%$, it appears that the calibration significantly improved the accuracy of the effort estimates. The default Pred (0.25) was 0.00%, and the calibrated Pred (0.25) was 83.33%. This also implies that the calibration greatly improved the estimating accuracy. The Wilcoxon signed-rank test indicated that the default model overestimated the software effort (Bias +), while the calibrated model had no over or underestimating bias (Unbiased). This

calibration was able to improve the estimate of five out of six records in the validation data set. This fact, along with the signs of the differences, indicates that this data set was quite homogeneous, and the calibration was primarily adjusting for the overestimation tendency of SoftCost-R with the default parameters.

Military Ground - Signal Processing. The military ground signal processing validation data set consisted of nine data records. The records and the results of this validation are given in Appendix D. The default effort MRE ranged from 16.70% to 148.71%, and the calibrated MRE ranged from 1.40% to 65.07%. The default effort MMRE was 42.98%, and the calibrated MMRE was 28.24%. The default effort RRMS was 61.19%, and the calibrated RRMS was 63.39%. Since the calibrated MRE range was less varied and closer to the ideal of 0%, the calibrated MMRE was much closer to the ideal of $\leq 25\%$, and the calibrated RRMS was almost unchanged, it appears that the calibration improved the accuracy of the effort estimates. The default Pred (0.25) was 11.11%, and the calibrated Pred (0.25) was 44.44%. Although the calibrated Pred (0.25) did not meet Conte's criteria, it was an improvement over the default value. This also implies that the calibration may have improved the estimating accuracy. The Wilcoxon signed-rank test indicated that both the default model and the calibrated model were unbiased. This calibration was able to improve the estimate of five out of nine records in the validation data set. This fact, along with the signs of the differences, indicates that the calibration was primarily adjusting the estimates slightly closer to the actual values by lowering the estimates.

Unmanned Space. The unmanned space validation data sets consisted of five U.S. data records and seven European data records, respectively. The records and the results of this validation are given in Appendix D. For the U.S. records, the default effort MRE ranged from 12.43% to 91.27%, and the calibrated MRE ranged from 19.47% to

85.27%. The U.S. default effort MMRE was 55.74%, and the calibrated MMRE was 47.97%. The U.S. default effort RRMS was 104.78%, and the calibrated RRMS was 92.28%. Since the calibrated MRE range was virtually unchanged, the calibrated MMRE improved only slightly, and the calibrated RRMS improved only slightly, it appears that the calibration had only a minor effect on the accuracy of the U.S. effort estimates. The U.S. default Pred (0.25) was 20.00%, and the calibrated Pred (0.25) was 20.00%. This also implies that the calibration did not significantly improve the estimating accuracy for the U.S. records. The U.S. Wilcoxon signed-rank test indicated that the default model was unbiased, while the calibrated model underestimated the effort (Bias -). This calibration was able to improve the estimate of four out of five records in the U.S. validation data set. This fact, along with the signs of the differences, indicates that the U.S. data set was homogeneous, and the calibration was primarily adjusting for a slight underestimation tendency of SoftCost-R with the default parameters. The European default effort MRE ranged from 20.88% to 457.57%, and the calibrated MRE ranged from 8.81% to 320.61%. The European default effort MMRE was 179.34%, and the calibrated MMRE was 127.28%. The European default effort RRMS was 78.98%, and the calibrated RRMS was 83.98%. Since the calibrated MRE range was less varied and closer to the ideal of 0%, the calibrated MMRE was closer to the ideal of $\leq 25\%$, and the calibrated RRMS was almost unchanged, it appears that the calibration improved the accuracy of the European effort estimates. The European default Pred (0.25) was 14.29%, and the calibrated Pred (0.25) was 14.29%. This implies that the calibration did not affect the European estimating accuracy. The European Wilcoxon signed-rank test indicated that both the default model and the calibrated model were unbiased. Note that the calibration was able to improve the estimate of five out of seven records in the European validation data set. This fact, along with the signs of the differences, indicates

that the European data set was relatively homogeneous, and the calibration was primarily adjusting for a slight overestimation tendency of SoftCost-R with the default parameters.

Ground in Support of Space. The ground in support of space validation data sets consisted of 15 U.S. data records and 30 European data records, respectively. The records and the results of this validation are given in Appendix D. For the U.S. records, the default effort MRE ranged from 6.07% to 995.22%, and the calibrated MRE ranged from 1.48% to 685.39%. The U.S. default effort MMRE was 273.43%, and the calibrated MMRE was 180.23%. The U.S. default effort RRMS was 312.53%, and the calibrated RRMS was 196.60%. Since the calibrated MRE range was less varied and closer to the ideal of 0%, the calibrated MMRE improved, and the calibrated RRMS improved, it appears that the calibration improved the accuracy of the U.S. effort estimates. The U.S. default Pred (0.25) was 13.33%, and the calibrated Pred (0.25) was 20.00%. This also implies that the calibration slightly improved the estimating accuracy for the U.S. records. The U.S. Wilcoxon signed-rank test indicated that both the default model and the calibrated model overestimated the effort (Bias +). This calibration was able to improve the estimate of 12 out of 15 records in the U.S. validation data set. This fact, along with the signs of the differences, indicates that the U.S. data set was somewhat homogeneous, and the calibration was primarily adjusting for an overestimation tendency of SoftCost-R with the default parameters. Although the model still overestimated after calibration, it did not overestimate as severely. The European default effort MRE ranged from 7.03% to 921.60%, and the calibrated MRE ranged from 5.12% to 197.53%. The European default effort MMRE was 304.99%, and the calibrated MMRE was 66.47%. The European default effort RRMS was 361.28%, and the calibrated RRMS was 83.65%. Since the calibrated MRE range was less varied and much closer to the ideal of 0%, the calibrated MMRE was closer to the ideal of $\leq$ 25%, and the calibrated RRMS was closer

to the ideal of ≤ 25%, it appears that the calibration significantly improved the accuracy of the European effort estimates. The European default Pred (0.25) was 20.00%, and the calibrated Pred (0.25) was 40.00%. This also implies that the calibration improved the European estimating accuracy. The European Wilcoxon signed-rank test indicated that the default model overestimated (Bias +), while the calibrated model was unbiased. This calibration was able to improve the estimate of 21 out of 25 records in the European validation data set. This fact, along with the signs of the differences, indicates that the European data set was relatively homogeneous, and the calibration was primarily adjusting for an overestimation tendency of SoftCost-R with the default parameters.

Military Mobile. The military mobile validation data set consisted of five data records. The records and the results of this validation are given in Appendix D. The default effort MRE ranged from 19.26% to 185.17%, and the calibrated MRE ranged from 8.31% to 99.50%. The default effort MMRE was 63.45%, and the calibrated MMRE was 41.95%. The default effort RRMS was 51.38%, and the calibrated RRMS was 39.46%. Since the calibrated MRE range was much less varied and closer to the ideal of 0%, the calibrated MMRE was closer to the ideal of ≤ 25%, and the calibrated RRMS was closer to the ideal of ≤ 25%, it appears that the calibration improved the accuracy of the effort estimates. The default Pred (0.25) was 20.00%, and the calibrated Pred (0.25) was 40.00%. Although the calibrated Pred (0.25) did not meet Conte's criteria, it was an improvement over the default value. This also implies that the calibration may have improved the estimating accuracy. The Wilcoxon signed-rank test indicated that both the default model and the calibrated model were unbiased. This calibration was able to improve the estimate of four out of five records in the validation data set. This fact, along with the signs of the differences, indicates that data set was relatively homogeneous, and

the calibration was primarily adjusting for a slight underestimation in the default model by increasing the estimates.

Missile. There were not enough suitable data records in this category to perform validation.

Military Specification Avionics. The military specification avionics validation data set consisted of five data records. The records and the results of this validation are given in Appendix D. The default effort MRE ranged from 10.21% to 179.70%, and the calibrated MRE ranged from 12.02% to 241.31%. The default effort MMRE was 71.27%, and the calibrated MMRE was 84.55%. The default effort RRMS was 75.77%, and the calibrated RRMS was 56.80%. Since the calibrated MRE range was more varied and further from the ideal of 0%, the calibrated MMRE was slightly further from the ideal of $\leq 25\%$, and the calibrated RRMS was only slightly closer to the ideal of $\leq 25\%$, it appears that the calibration did not improve the accuracy of the effort estimates and may have made them slightly worse. The default Pred (0.25) was 20.00%, and the calibrated Pred (0.25) was 20.00%. This also implies that the calibration did not affect the estimating accuracy. The Wilcoxon signed-rank test indicated that both the default model and the calibrated model were unbiased. This calibration was able to improve the estimate of three out of five records in the validation data set. This fact, along with the signs of the differences, indicates that the calibration was primarily adjusting the estimates slightly closer to the actual values by lowering the estimates.

Validation Summary. The intent of this section was to present the results of the validation data set estimates and to present analysis and observations pertaining to these results. A summary of these results and statistics is given in Table 12. Conclusions and implications derived from these results will be presented in Chapter V.

Table 12. Validation Results Summary

Environment and Application Type	Default MRE Range (%)	Calibrated MRE Range (%)	Default MMRE (%)	Calibrated MMRE (%)	Default RRMS (%)	Calibrated RRMS (%)	Default Pred (0.25) (%)	Calibrated Pred (0.25) (%)	Default Wilcoxon Signed-Rank Test	Calibrated Wilcoxon Signed-Rank Test
Mil Ground - C&C	22.48-443.09	1.60-128.39	189.52	51.86	343.33	87.04	16.67	50.00	Bias +	Unbiased
Mil Ground - SP	16.70-148.71	1.40-65.07	42.98	28.24	61.19	63.39	11.11	44.44	Unbiased	Unbiased
Unmanned Space - U.S.	12.43-91.27	19.47-85.27	55.74	47.97	104.78	92.28	20.00	20.00	Unbiased	Bias -
Unmanned Space - European	20.88-457.57	8.81-320.61	179.34	127.28	78.98	83.98	14.29	14.29	Unbiased	Unbiased
Ground in Support of Space - U.S.	6.07-995.22	1.48-685.39	273.43	180.23	312.53	196.60	13.33	20.00	Bias +	Bias +
Ground in Support of Space - European	7.03-921.60	5.12-197.53	304.99	66.47	361.28	83.65	8.00	36.00	Bias +	Unbiased
Military Mobile	19.26-185.17	8.31-99.50	63.45	41.95	51.38	39.46	20.00	40.00	Unbiased	Unbiased
Missile	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Mil-Spec Avionics	10.21-179.70	12.02-241.31	71.27	84.55	75.77	56.80	20.00	20.00	Unbiased	Unbiased

V. Conclusions and Recommendations

Overview

The objectives of this research were to calculate new AWFFAC and KSLECPM parameters for the SoftCost-R model by calibrating it to the SMC SWDB and to validate whether or not this calibration method improved SoftCost-R's estimating accuracy. If valid, the calibrated AWFFAC and KSLECPM parameters would offer an improved estimating capability for software development projects that are similar to those in the calibration data set, i.e., that are in the same environment and application type areas. The calibration method itself would offer a means to update these parameters for these and other homogeneous data sets in the future as more records are added to the SMC SWDB. This chapter gives conclusions based on this calibration and validation It also gives recommendations on the use and calibration of the SoftCost-R model and on possible future research in this area. The research questions proposed in Chapter I, along with brief answers, are included here:

1. What is the uncalibrated accuracy of the SoftCost-R model when estimating efforts in the SMC SWDB? The uncalibrated SoftCost-R model was not very accurate. It only estimated approximately 30% of the projects within 50% of their actual results. This was primarily due to a conservative overestimation bias.

2. Can SoftCost-R be calibrated to subsets of the SMC SWDB? Yes, SoftCost-R can be calibrated by changing the AWFFAC and KSLECPM values using a best fit least squares linear regression method.

3. What is the accuracy of the SoftCost-R model after it has been calibrated to the SMC SWDB? The calibrated SoftCost-R model was better than the uncalibrated model, but it still did not achieve the desired accuracy. It only estimated approximately 47% of the projects within 50% of their actual results.

4. What improvement was achieved due to calibration? The calibration corrected the conservative overestimating bias and improved the accuracy by approximately a factor of two (see Table 13). The MMRE and RRMS

percentages were less than half of the uncalibrated percentages. The Pred (0.25) was more than double the uncalibrated value, and the Pred (0.50) was approximately 50% greater than the uncalibrated result.

Conclusions

Overall, the calibration of SoftCost-R to the SMC SWDB appears to have been successful. However, the success is neither complete nor all encompassing, since none of the environments and application types achieved the goodness of fit goals and some of them did not experience significant accuracy improvements from the calibration. The results of the calibration, validation, and statistical analysis were given in Chapter IV and will not be repeated here. Instead, this section will offer general conclusions based on those results. Table 13 gives a summary of the key validation statistics, calculated using a weighted average across all of the data sets, and also adds the statistic Pred (0.50), which is the percentage of time that the estimate was within 50% of the actual effort. This last statistic was added in order to evaluate the SoftCost-R user's manual claim that SoftCost-R estimates were accurate to within 50% of the actual effort 67% of the time (Reifer Consultants, 1989:U-82). In the case of the SMC SWDB, the calibrated estimates came close to achieving this accuracy, but the uncalibrated estimates for which the claim was made were less that half this accurate. It should also be noted that the default Wilcoxon signed-rank bias in favor of overestimation is probably due to the fact that the SoftCost-R model gives an estimate that has a 70% confidence level, rather than the most likely 50% confidence level estimate that other models report (Reifer Consultants, 1989:U-82). The confidence level is defined as the probability that the project will be completed for that amount of effort or less. The calibration cures this overestimation and brings the estimate back down to a 50% confidence level.

Table 13. Weighted Average Statistics for All Validation Data Sets

Statistic	Default Weighted Average	Calibrated Weighted Average	Improvement Due to the Calibration
MRE Range	6.07-995.22%	1.48-685.39%	4.59-309.83%
MMRE	200.75%	86.93%	113.82%
RRMS	234.33%	99.53%	134.80%
Pred (0.25)	12.99%	31.17%	18.18%
Pred (0.50)	29.87%	46.75%	16.88%
Wilcoxon Signed-Rank Test	Bias + (Overestimate)	Unbiased	Eliminates Overestimation

Of greater concern was the inability of the calibration to achieve a MMRE $\leq$ 25%, a RRMS $\leq$ 25%, or a Pred (0.25) $\geq$ 75% level of accuracy for the calibration data sets that were the basis of the calibration. This indicates one or more of the following may be true:

1. The data sets are not homogeneous enough or large enough to achieve accurate specific predictions, i.e., the sample population is too varied or too small to be statistically significant. This could only be corrected by larger and more homogeneous data sets or a model with factors that account for more of the variation.

2. The lack of information in the data fields of the SMC SWDB, due to data records that leave these fields blank and also to the fact that several of the SoftCost-R inputs did not have corresponding data fields in the SMC SWDB, did not allow the model to account for the variation in the data. This could be corrected by adding more fields to the SMC SWDB to cover the SoftCost-R input factors, and by encouraging organizations that submit records to the SMC SWDB to provide accurate inputs to all of the fields. Without data records that are complete, i.e., have information in all of the applicable data fields, it becomes a question of data quantity versus quality. That is, a database of over 2600 records, most of which are incomplete, is less useful than a much smaller database of complete and homogeneous records.

3. The wrong parameters were calibrated. Improved accuracy may be achieved by calibrating other parameters within the SoftCost-R model equations. The most promising would be Gamma_W, which is a linear multiplier of the entire equation, and A, which is the exponent to which the effective size is raised in the equation. There may be more variability in the data than implied by the few default values for Gamma_W. A wider range and more categories for Gamma_W may account for more of this variability. It is also quite likely that

the size is not a linear factor in determining software effort. Therefore, the fact that A defaults to 1.0 may not be reasonable. Similar to the COCOMO model, this exponent may need to be calibrated to the different data sets, or at least to the database as a whole.

4. The SoftCost-R model is not capable of achieving this level of prediction accuracy. The only solution to this problem would be to use another cost estimating model or to update the SoftCost-R model by adding more input factors to account for more of the variability.

However, due to the large number of SoftCost-R factor inputs covering a broad spectrum of software development characteristics, it is believed that the answer to this problem lies primarily with items 1, 2, and 3. The first two items, especially item 2, may be solved by calibrating SoftCost-R to organizational databases which are both complete and homogeneous. This should provide much more accurate estimates for future software developments in that organization.

A subjective summary of the overall goodness of fit of the calibrated SoftCost-R model and the accuracy improvement due to the calibration is given in Table 14. This shows that calibration provided the most benefit for the command and control and European ground in support of space categories, while providing virtually no improvement for the U.S. unmanned space and military specification avionics categories. After calibration, the model was nearly successful at obtaining a good fit for the command and control, signal processing, U.S. unmanned space, and military mobile categories. These categories achieved an MMRE $\leq$ 50%, a RRMS $\leq$ 50%, or a Pred (0.25) $\geq$ 50%. None of the other categories even came close to a good fit. It was especially interesting to note that both of the European data sets experienced larger accuracy improvements than their U.S. counterparts. This may be due to the fact that in both cases they had worse default accuracy to begin with. This poor default accuracy may be due to software development standards, techniques, and practices that are different in Europe than the methods assumed as a basis for the SoftCost-R model.

Table 14. Calibration Effectiveness Based on Validation Results

Environment and Application Type	Satisfaction of Goodness of Fit Criteria	Improvement Due to the Calibration
Mil Ground - C&C	Nearly Successful	Large Improvement
Mil Ground - SP	Nearly Successful	Improvement
Unmanned Space - U.S.	Nearly Successful	Neutral
Unmanned Space - European	Not Successful	Minor Improvement
Ground in Support of Space - U.S.	Not Successful	Improvement
Ground in Support of Space - European	Not Successful	Large Improvement
Military Mobile	Nearly Successful	Improvement
Missile	N/A	N/A
Mil-Spec Avionics	Not Successful	Neutral

Recommendations

This section will list several recommendations concerning calibration of the SoftCost-R model and future research.

When calibrating SoftCost-R in the future, it is recommended that the calibration be made to more homogeneous and complete data sets in order to achieve greater calibrated accuracy. The importance of complete and accurate historical data cannot be overstated, since historical data are the foundation for algorithmic models and regression based (best fit) calibration methods. This could be accomplished by calibrating to a database that is specific to a single organization and contains valid inputs for all of the SoftCost-R factors. Calibration data sets could also be stratified by language, as well as environment and application type. Although this may result in smaller data sets, they will be much more homogeneous, and greater accuracy should be achieved. The current calibration method that does not take into account programming language covers too broad of a spectrum from machine and assembly code routines to Very High Order

Languages. In order to achieve this improved accuracy, there must be a strong dedication to collection and reporting of software development data. This data must be accurate and complete. Also, if at all possible, a method should be developed to allow calibration efforts to consider the development contractor, while still maintaining the contractor's anonymity and willingness to provide complete and accurate information.

The recommendations for future research fall into three categories:

1. Calibrate SoftCost-R using other calibrateable parameters, such as Gamma_W and A, in order to determine if more accurate estimates can be achieved. Use the same spreadsheet and solver approach. As computing power increases, more than two parameters could be calibrated simultaneously.

2. Re-calibrate SoftCost-R using the method from this research, but using the actual data record inputs for new, reused, and modified code size and for development phases included. Rather than using the normalized effective size and normalized effort, this original raw data could be entered into SoftCost-R using its software size submodel and the Life Cycle Coverage factor. If this resulted in more accurate estimates, then it would call into question the basis for the SMC SWDB normalization procedures, at least as they relate to SoftCost-R. Further analysis of these normalization procedures could then be performed.

3. Determine calibration procedures for other SoftCost family models, such as SoftCost-Ada and SoftCost-OO. These newer models have been updated more recently than SoftCost-R. Their internal equations are based on more recent software databases that are more representative of current software development efforts, especially in the Ada and object-oriented areas. Therefore, determining a calibration method for these models should be a priority. Even if a calibration method cannot be determined, research comparing the default estimates of these models to the calibrated estimates of SoftCost-R could be performed. If these default models are more accurate than a calibrated SoftCost-R model, then future research and use of SoftCost-R could be curtailed.

As a final comment, it should be noted that there were several discrepancies and missing pieces of information in the equations section of the SoftCost-R reference manual and one error in the source code that displays the project reports. These errors are essential knowledge to anyone trying to understand the equations underlying the

SoftCost-R model, especially if the model's equations are being emulated in another software package. The source code error in the project reports screen would even affect a more casual user of SoftCost-R. This was especially true for the input factor Adaptation Requirements. The numerical values for this factor for the various levels (low, normal, high, etc.) were different in all four different places in which they occur: reference manual, calibration file (softcost.cal), model project summary report display (the software error), and source code. Also, the numerical values for the factor Concurrent Hardware Development in the reference manual are inverted, since high concurrent development should be a positive (greater than nominal) multiplier to increase the effort and low concurrent development should be a negative (less than nominal) multiplier. Another non-standard convention in the documentation is the use of the C programming language term "log" when what is meant is more commonly referred to as "ln" (a logarithm to the natural base e). Also, although both the RCI company president and the SoftCost-R support programmer were eager and willing to provide assistance and interpretation of the documentation, some of the questions could not be answered immediately due to the loss of corporate knowledge about this model that occurred with the departure of Donald Reifer. This same problem may face anyone that calibrates other parameters in SoftCost-R or other SoftCost family models (SoftCost-Ada, SoftCost-OO).

Appendix A. Acronyms and Glossary of Terms

Term	Definition or Expansion
A	Calibrated Effort Exponent (a constant in the SoftCost-R equations)
A_1	Product of the nine Reifer cost drivers (a variable in the SoftCost-R equations)
A_2	Sum of the modified Tausworthe productivity adjustment factors (a variable in the SoftCost-R equations)
AE	Actual Effort (abbreviation used in the calibration equations)
AWF	Work Effort Tradeoff Factor (a variable in the SoftCost-R equations)
AWFFAC	Productivity adjustment constant = Average Work Force Factor (a calibrateable parameter in SoftCost-R)
CADJ	Calibrated Space Critical Adjustment Factor (a constant in the SoftCost-R equations)
COCOMO	Constructive Cost Model
DoD	Department of Defense
EADJ	Calibrated Easy Adjustment Factor (a constant in the SoftCost-R equations)
Effort_11	SoftCost-R source code terminology for W_N
FADJ	Calibrated Free Requirements Change Adjustment Factor (a constant in the SoftCost-R equations)
FULLUP	Calibrated Full Up Training Adjustment Factor (a constant in the SoftCost-R equations)
Gamma_W	Application Domain specific effort calibration constant (a constant in the SoftCost-R equations)
GAO	Government Accounting Office
HADJ	Calibrated Hard Adjustment Factor (a constant in the SoftCost-R equations)
Index	SoftCost-R source code terminology for A_2

Index_0	Calibrated ln-ratio sum (a constant in the SoftCost-R equations)
IPE	Initial Predicted Effort (abbreviation used in the calibration equations)
KSLEC	Thousands of Source Lines of Executable Code. (Same as KSLOC)
KSLECPM	Productivity ratio factor = thousands (K) of Source Lines of Executable Code Productivity Multiplier) (a calibrateable parameter in SoftCost-R)
KSLOC	Thousands of Source Lines Of Code
LADJ	Calibrated Assembly Language Adjustment Factor (a constant in the SoftCost-R equations)
MMRE	Mean Magnitude of Relative Error
MRE	Magnitude of Relative Error
NPE	New Predicted Effort (abbreviation used in the calibration equations)
P_η	SoftCost-R source code terminology for A_1
P_0	Productivity factor calibrated from calibration values (a variable in the SoftCost-R equations)
PA	Calculated Effort Technology Constant (a multiplier in the SoftCost-R equations)
PM	Person-Month = one person working for one month. Usually 152 or 160 person-hours.
Pred (k/n)	Prediction level = percentage of estimates within (100 * k/n)% of the actual costs
PRICE S	Programmed Review of Information for Costing and Evaluation Software
productivity	Source Lines of Code per Person-Month (SLOC/PM)
RCI	Resource Calculations Inc. - Owner of SoftCost-R
Reifer	Donald J. Reifer. SoftCost-R Version 8.4 incorporated extensions developed by Reifer to improve the SoftCost model
REVIC	Revised Enhanced Version of Intermediate COCOMO
RMS	Root Mean Square

RRMS	Relative Root Mean Square
SASET	Software Architecture Sizing and Estimation Tool
SEER-SEM	System Estimation and Evaluation of Resources - Software Estimation Model
SLIM	Software Life Cycle Model
SLOC	Source Lines of Code
SMC	Space and Missile Systems Center at Los Angeles Air Force Base
SMC SWDB	Software Database - A database of software development and support information collected by SMC. Currently it includes 2638 entries.
SMC/FMC	SMC Directorate of Cost - Developer and maintainer of the SMC SWDB
SoftCost-R	Software Costing - Real-time - a parametric model used to estimate the cost and schedule of object-oriented software development efforts
softcost.cal	The text file in SoftCost-R which contains the default calibration parameters and constants for the model.
SSCAG	Space Systems Cost Analysis Group - Major contributor to the SMC SWDB
SSM	Software Sizing Model
TADJ	Calibrated Time Critical Adjustment Factor (a constant in the SoftCost-R equations)
Tausworthe	Dr. Robert C. Tausworthe. Developed the SoftCost mathematical model for NASA's Jet Propulsion Laboratory in 1981
WBS	Work Breakdown Structure
W_C	Effort Normalization Constant (a constant in the SoftCost-R equations)
W_N	Effort Normalization Factor (a variable in the SoftCost-R equations)

Appendix B. SMC SWDB Field to SoftCost-R Factor Correspondence

SoftCost-R Input Factor	SMC SWDB Data Field and Conversion Method *
Type of Software	Entered manually - based on query type
System Architecture	Automatic - 4.6 System Architecture
User Involvement	
Organizational Interface Complexity	
Staff Resource Availability	
Computer Resource Availability	Automatic - 4.23.9 Resource Dedication and 4.23.10 Resource/Support Location
Security Requirements	Automatic - 4.3.7 Security Level
Concurrent Hardware Development	Automatic - 4.27 New Hardware Design, 4.28 Hardware Developed in Parallel with Software, and 4.3.4 Re-hosting Requirements
Code Delivery Requirements	
Degree of Standardization	Entered manually - based on 2.9 Development Standard (automatic default if not updated manually)
Life Cycle Coverage	Automatic default based on the phases included in the SMC SWDB effective effort
Scope of Support	
Use of Modern Software Methods	Automatic - 4.23.13 Modern Practices Experience
Use of Peer Reviews	
Use of Software Tools/Environment	Automatic - 4.23.14 Automated Tool Support
Software Tool/Environment Stability	Automatic - 4.23.4 Development System Volatility
Geographical Co-Location	Automatic - 4.23.8 Multiple Site Development
Program Complexity	Automatic - 4.3.1 Application Complexity
Database Complexity	
Requirements Complexity	Automatic - 4.23.5 Specification Level
Requirements Volatility	Automatic - 4.3.3 Requirements Volatility
Degree of Optimization	Automatic - 4.3.8 Memory Constraints and 4.3.9 Timing Constraints
Degree of Real-Time	Automatic - 4.3.10 Real Time
Adaptation Requirements	Automatic - 4.3.4 Re-hosting Requirements
Database Size	Automatic - 3.4 Total Database Size (in bytes) and the normalized effective size figure
Analyst Capability	Automatic - 4.8.2 Personnel Capabilities
Programmer Capability	Automatic - 4.8.2 Personnel Capabilities
Applications Experience	Automatic - 4.8.1 Personnel Experience

Language Experience	Automatic - 4.8.5 Team Programming Language Experience
Environment Experience	Automatic - 4.8.7 Development System Experience
Methodology Experience	Automatic - 4.8.6 Development Methods Experience
Customer Experience	
Team Capability	Automatic - 4.8.2 Personnel Capabilities

* All factors that do not have a corresponding field in the SMC SWDB or for which the field in the SMC SWDB is empty are based on default values for the SoftCost-R quick run type, which is determined by the SMC SWDB query type.

(Reifer Consultants, Inc., 1989:R-19 to R-82; Management and Consulting Research, Inc. and Cost Management Systems, Inc., 1995:B-7 to B-29)

Appendix C. Calibration Data Effort Estimates and Statistics

Command and Control Calibration Records - U.S.

Record #	Effective Size (SLOC)	Normalized Actual Effort (PM)	Default Effort (PM)	Calibrated Effort (PM)	Default MRE (%)	Calibrated MRE (%)
7	45,057	120	457	201	280.83	67.13
50	144,000	684	1487	645	117.40	5.76
120	25,842	95	279	117	193.57	23.46
124	23,881	139	258	108	85.42	22.03
152	69,772	286	753	317	163.28	10.72
2510	43,437	172	313	174	82.00	1.01
Default Data Set Pred (0.25) = 0.00% Calibration Data Set Pred (0.25) = 83.33% Calibrated AWFFAC = 0.676 Calibrated KSLECPM = 0.730						

Signal Processing Calibration Records - U.S.

Record #	Effective Size (SLOC)	Normalized Actual Effort (PM)	Default Effort (PM)	Calibrated Effort (PM)	Default MRE (%)	Calibrated MRE (%)
54	45,035	127	470	243	269.79	91.34
127	16,016	13	137	74	954.06	468.26
132	46,595	278	399	215	43.40	22.69
133	123,710	645	1058	571	64.10	11.53
135	23,787	264	204	110	22.91	58.44
140	70,020	6	599	323	9884.50	5282.81
142	28,782	348	246	133	29.24	61.85
143	23,703	86	203	109	135.81	27.13
153	11,534	149	99	53	33.77	64.29
154	8,965	109	77	41	29.63	62.06
Default Data Set Pred (0.25) = 10.00% Calibration Data Set Pred (0.25) = 30.00% Calibrated AWFFAC = 1.712 Calibrated KSLECPM = 0.685						

Unmanned Space Calibration Records - U.S.

Record #	Effective Size (SLOC)	Normalized Actual Effort (PM)	Default Effort (PM)	Calibrated Effort (PM)	Default MRE (%)	Calibrated MRE (%)
3	80,000	583	1273	2148	118.27	268.36
84	6,000	796	95	161	88.04	79.82
85	1,950	204	31	52	84.80	74.34
2623	16,759	305	260	398	14.90	30.37
2624	16,759	305	260	398	14.90	30.37
2625	16,759	305	260	398	14.90	30.37
Default Data Set Pred (0.25) = 60.00% Calibration Data Set Pred (0.25) = 0.00% Calibrated AWFFAC = 10.00 Calibrated KSLECPM = 0.881						

Unmanned Space Calibration Records - European

Record #	Effective Size (SLOC)	Normalized Actual Effort (PM)	Default Effort (PM)	Calibrated Effort (PM)	Default MRE (%)	Calibrated MRE (%)
2566	5,000	48	77	58	61.34	21.70
2567	13,000	131	201	152	53.70	15.94
2570	14,000	66	217	164	228.54	147.84
2572	12,000	28	186	140	563.78	400.73
2576	11,000	202	170	129	15.66	36.38
2578	5,000	63	77	58	22.92	7.27
2580	7,000	93	108	82	16.58	12.06
2582	15,000	313	232	175	25.78	44.01
Default Data Set Pred (0.25) = 37.50% Calibration Data Set Pred (0.25) = 50.00% Calibrated AWFFAC = 1.965 Calibrated KSLECPM = 0.471						

Ground in Support of Space Calibration Records - U.S.

Record #	Effective Size (SLOC)	Normalized Actual Effort (PM)	Default Effort (PM)	Calibrated Effort (PM)	Default MRE (%)	Calibrated MRE (%)
75	116,800	912	1366	1018	49.73	11.66
76	14,000	115	141	100	22.87	12.88
77	56,200	523	657	490	25.63	6.32
78	48,300	478	552	409	15.53	14.50
80	69,450	296	794	588	168.26	98.53
81	22,900	164	262	194	59.65	18.15
82	16,300	140	186	138	33.12	1.48
83	6,800	57	78	58	36.40	0.94
91	52,275	1169	545	391	53.36	66.55
105	21,000	5	219	157	4280.88	3041.56
107	8,000	160	83	60	47.85	62.60
115	13,000	109	136	97	24.40	10.79
117	66,843	652	697	500	6.94	23.32
329	34,650	57	305	207	435.93	262.77
332	60,087	70	509	340	626.91	385.30
Default Data Set Pred (0.25) = 26.67% Calibration Data Set Pred (0.25) = 60.00% Calibrated AWFFAC = 2.901 Calibrated KSLECPM = 0.553						

Ground in Support of Space Calibration Records - European

Record #	Effective Size (SLOC)	Normalized Actual Effort (PM)	Default Effort (PM)	Calibrated Effort (PM)	Default MRE (%)	Calibrated MRE (%)
2529	45,000	90	460	146	410.80	62.29
2532	126,000	244	1287	409	427.55	67.61
2533	16,000	18	163	52	808.09	188.52
2534	6,000	9	61	19	581.07	116.39
2536	22,000	636	225	71	64.66	88.77
2539	84,000	793	585	273	8.22	65.62
2542	6,000	56	61	19	9.46	65.22
2544	22,000	118	225	71	90.47	39.48
2548	150,000	222	1532	487	590.27	119.31
2549	21,000	43	215	68	398.92	58.52
2552	12,000	30	123	39	308.64	29.83
2553	35,000	85	358	114	320.66	33.65
2557	11,000	15	112	36	649.18	138.03
2560	18,000	145	184	58	26.82	59.71
2564	50,000	278	511	162	83.74	41.62
2583	62,000	497	633	201	27.44	59.51
2585	14,000	23	143	45	521.85	97.57
2587	32,000	72	327	104	354.05	44.26
2589	10,000	140	102	32	27.03	76.82
2596	40,000	221	409	130	84.91	41.25
2597	75,000	130	766	243	489.39	87.26
2599	49,000	526	501	159	4.83	69.76
2601	80,000	197	817	260	314.86	31.81
2602	50,000	138	511	162	270.15	17.60
2605	12,000	36	123	39	240.53	8.19

Default Data Set Pred (0.25) = 12.00%
Calibration Data Set Pred (0.25) = 8.00%
Calibrated AWFFAC = 1.597
Calibrated KSLECPM = 1.095

Military Mobile Calibration Records - U.S.

Record #	Effective Size (SLOC)	Normalized Actual Effort (PM)	Default Effort (PM)	Calibrated Effort (PM)	Default MRE (%)	Calibrated MRE (%)
34	17,134	83	179	110	115.32	32.16
303	30,000	237	334	244	40.96	2.76
2505	7,448	180	105	142	41.73	21.21
2506	6,317	152	91	131	40.06	13.73
2508	58,789	1418	884	1423	37.62	0.35
Default Data Set Pred (0.25) = 0.00% Calibration Data Set Pred (0.25) = 80.00% Calibrated AWFFAC = 7.842 Calibrated KSLECPM = 1.083						

Military Specification Avionics Calibration Records - U.S.

Record #	Effective Size (SLOC)	Normalized Actual Effort (PM)	Default Effort (PM)	Calibrated Effort (PM)	Default MRE (%)	Calibrated MRE (%)
10	43,207	370	734	492	98.43	32.86
302	45,353	400	703	516	75.76	28.99
346	40,000	654	525	455	19.72	30.42
2512	33,158	245	603	377	146.03	53.97
2618	26,000	363	485	296	33.65	18.51
Default Data Set Pred (0.25) = 20.00% Calibration Data Set Pred (0.25) = 20.00% Calibrated AWFFAC = 0..000 Calibrated KSLECPM = 0.349						

Appendix D. Validation Data Effort Estimates and Statistics

Command and Control

Record #	Effective Size (KSLOC)	Actual Effort (PM)	Default Predicted Effort (PM)	Calibrated Predicted Effort (PM)	Default Difference (PM)	Calibrated Difference (PM)	Default MRE (%)	Calibrated MRE (%)	Meets Default Pred(0.25) ≥ 75%	Meets Calibrated Pred(0.25) ≥ 75%
9	128	517	1512	598	995	81	192.49	15.71	N	Y
141	162	322	1749	735	1427	413	443.09	128.39	N	N
145	19	101	200	84	99	-17	98.32	16.60	N	Y
150	22	100	234	98	134	-2	133.99	1.60	N	Y
155	8	74	91	38	17	-36	22.48	48.49	Y	N
2517	85	167	579	335	412	168	246.77	100.37	N	N

Default MMRE = 189.52%
Calibrated MMRE = 51.86%
Default RMS = 733.02
Calibrated RMS = 185.82
Default RRMS = 343.33%
Calibrated RRMS = 87.04%
Default Pred (0.25) = 16.67%
Calibrated Pred (0.25) = 50.00%

Signal Processing

Record #	Effective Size (KSLOC)	Actual Effort (PM)	Default Predicted Effort (PM)	Calibrated Predicted Effort (PM)	Default Difference (PM)	Calibrated Difference (PM)	Default MRE (%)	Calibrated MRE (%)	Meets Default Pred(0.25) ≥ 75%	Meets Calibrated Pred(0.25) ≥ 75%
126	48	165	410	221	245	56	148.71	34.08	N	N
130	72	738	615	331	-123	-407	16.70	55.09	Y	N
131	29	192	249	134	57	-58	29.88	29.98	N	N
134	45	228	381	205	153	-23	67.09	9.92	N	Y
136	12	154	104	56	-50	-98	32.66	63.70	N	N
137	60	274	515	278	241	4	88.08	1.40	N	Y
138	14	190	123	66	-67	-124	35.21	65.07	N	N
144	30	145	255	137	110	-8	75.85	5.20	N	Y
147	32	192	271	146	79	-46	41.35	23.80	N	Y

Default MMRE = 42.98%
Calibrated MMRE = 28.24%
Default RMS = 143.66
Calibrated RMS = 148.83
Default RRMS = 61.19%
Calibrated RRMS = 63.39%
Default Pred (0.25) = 11.11%
Calibrated Pred (0.25) = 44.44%

Unmanned Space - U.S.

Record #	Effective Size (KSLOC)	Actual Effort (PM)	Default Predicted Effort (PM)	Calibrated Predicted Effort (PM)	Default Difference (PM)	Calibrated Difference (PM)	Default MRE (%)	Calibrated MRE (%)	Meets Default Pred(0.25) ≥ 75%	Meets Calibrated Pred(0.25) ≥ 75%
86	6	200	95	161	-105	-39	52.28	19.47	N	Y
112	8	1511	132	223	-1379	-1288	91.27	85.27	N	N
113	20	1248	310	523	-938	-725	75.15	58.06	N	N
306	9	94	106	52	12	-42	12.43	44.75	Y	N
2622	20	558	293	378	-265	-180	47.55	32.29	N	N

Default MMRE = 55.74%
Calibrated MMRE = 47.97%
Default RMS = 756.70
Calibrated RMS = 666.46
Default RRMS = 104.78%
Calibrated RRMS = 92.28%
Default Pred (0.25) = 20.00%
Calibrated Pred (0.25) = 20.00%

Unmanned Space - European

Record #	Effective Size (KSLOC)	Actual Effort (PM)	Default Predicted Effort (PM)	Calibrated Predicted Effort (PM)	Default Difference (PM)	Calibrated Difference (PM)	Default MRE (%)	Calibrated MRE (%)	Meets Default Pred(0.25) ≥ 75%	Meets Calibrated Pred(0.25) ≥ 75%
2574	34	181	527	397	346	216	190.94	119.47	N	N
2575	9	25	139	105	114	80	457.57	320.61	N	N
2577	22	768	341	257	-427	-511	55.63	66.53	N	N
2579	32	410	496	374	86	-36	20.88	8.81	Y	Y
2581	30	764	465	351	-299	-413	39.18	54.12	N	N
2598	14	45	217	164	172	119	381.85	263.49	N	N
2607	5	37	77	58	40	21	109.30	57.89	N	N

Default MMRE = 179.34%
Calibrated MMRE = 127.28%
Default RMS = 251.62
Calibrated RMS = 267.55
Default RRMS = 78.98%
Calibrated RRMS = 83.98%
Default Pred (0.25) = 14.29%
Calibrated Pred (0.25) = 14.29%

Ground in Support of Space - U.S.

Record #	Effective Size (KSLOC)	Actual Effort (PM)	Default Predicted Effort (PM)	Calibrated Predicted Effort (PM)	Default Difference (PM)	Calibrated Difference (PM)	Default MRE (%)	Calibrated MRE (%)	Meets Default Pred(0.25) ≥ 75%	Meets Calibrated Pred(0.25) ≥ 75%
74	12	80	134	99	54	19	67.22	23.75	N	Y
79	50	432	575	426	143	-6	33.13	1.48	N	Y
88	117	244	1220	875	976	631	400.16	258.67	N	N
89	225	602	2347	1683	1745	1081	289.85	179.56	N	N
90	95	1055	991	711	-64	-344	6.07	32.65	Y	N
93	250	401	2608	1870	2207	1469	550.29	366.33	N	N
97	80	530	834	598	304	68	57.44	12.90	N	Y
98	90	86	942	675	856	589	995.22	685.39	N	N
99	8	234	83	60	-151	-174	64.34	74.43	N	N
106	16	206	170	122	-36	-84	17.47	40.81	Y	N
114	163	235	1700	1219	1465	984	623.24	418.64	N	N
116	400	1468	4168	2989	2700	1521	183.95	103.63	N	N
118	358	765	3734	2678	2969	1913	388.13	250.04	N	N
119	278	787	2905	2083	2118	1296	269.10	164.68	N	N
331	7	18	58	39	40	21	223.09	114.30	N	N

Default MMRE = 273.43%
Calibrated MMRE = 180.23%
Default RMS = 1,471.61
Calibrated RMS = 925.72
Default RRMS = 312.53%
Calibrated RRMS = 196.60%
Default Pred (0.25) = 13.33%
Calibrated Pred (0.25) = 20.00%

Ground in Support of Space - European

Record #	Effective Size (KSLOC)	Actual Effort (PM)	Default Predicted Effort (PM)	Calibrated Predicted Effort (PM)	Default Difference (PM)	Calibrated Difference (PM)	Default MRE (%)	Calibrated MRE (%)	Meets Default Pred(0.25) ≥ 75%	Meets Calibrated Pred(0.25) ≥ 75%
2531	130	345	1328	422	983	77	284.95	22.31	N	Y
2540	18	74	184	58	110	-16	148.50	21.05	N	Y
2543	11	105	112	36	7	-69	7.03	66.00	Y	N
2545	19	42	194	62	152	20	362.15	46.84	N	N
2546	42	85	429	136	344	51	404.79	60.38	N	N
2547	100	100	1022	325	922	225	921.60	224.58	N	N
2550	24	89	245	78	156	-11	175.49	12.47	N	Y
2551	19	65	194	62	129	-3	198.62	5.12	N	Y
2554	24	31	245	78	214	47	690.92	151.29	N	N
2555	83	103	848	269	745	166	723.23	161.56	N	N
2556	11	12	112	36	100	24	836.47	197.53	N	N
2558	55	292	562	179	270	-113	92.43	38.86	N	N
2561	47	331	480	153	149	-178	45.06	53.91	N	N
2562	29	234	296	94	62	-140	26.61	59.77	N	N
2563	17	196	174	55	-22	-141	11.39	71.85	Y	N
2584	7	12	72	23	60	11	495.94	89.34	N	N
2586	100	186	1022	325	836	139	449.25	74.51	N	N
2588	35	128	358	114	230	-14	179.34	11.25	N	Y
2590	16	59	163	52	104	-7	177.05	11.98	N	Y
2591	10	55	102	32	47	-23	85.75	40.98	N	N
2594	45	156	460	146	304	-10	194.69	6.37	N	Y
2595	14	58	143	45	85	-13	146.59	21.65	N	Y
2603	55	225	562	179	337	-46	149.73	20.66	N	Y
2608	55	71	562	179	491	108	691.38	151.44	N	N
2609	30	60	306	97	246	37	410.80	62.29	N	N

Default MMRE = 304.99%
Calibrated MMRE = 66.47%
Default RMS = 400.16
Calibrated RMS = 92.65
Default RRMS = 361.28%
Calibrated RRMS = 83.65%
Default Pred (0.25) = 8.00%
Calibrated Pred (0.25) = 36.00%

Military Mobile

Record #	Effective Size (KSLOC)	Actual Effort (PM)	Default Predicted Effort (PM)	Calibrated Predicted Effort (PM)	Default Difference (PM)	Calibrated Difference (PM)	Default MRE (%)	Calibrated MRE (%)	Meets Default Pred(0.25) ≥ 75%	Meets Calibrated Pred(0.25) ≥ 75%
348	18	396	320	790	-76	394	19.26	99.50	Y	N
2456	63	221	630	343	409	122	185.17	55.15	N	N
2502	26	633	375	529	-258	-104	40.70	16.44	N	Y
2503	32	783	476	718	-307	-65	39.15	8.31	N	Y
2507	27	647	434	843	-213	196	32.99	30.32	N	N

Default MMRE = 63.45%
Calibrated MMRE = 41.95%
Default RMS = 275.39
Calibrated RMS = 211.50
Default RRMS = 51.38%
Calibrated RRMS = 39.46%
Default Pred (0.25) = 20.00%
Calibrated Pred (0.25) = 40.00%

Military Avionics

Record #	Effective Size (KSLOC)	Actual Effort (PM)	Default Predicted Effort (PM)	Calibrated Predicted Effort (PM)	Default Difference (PM)	Calibrated Difference (PM)	Default MRE (%)	Calibrated MRE (%)	Meets Default Pred(0.25) ≥ 75%	Meets Calibrated Pred(0.25) ≥ 75%
11	33	198	539	374	341	176	172.12	88.91	N	N
12	22	112	313	251	201	139	179.70	123.75	N	N
13	58	752	988	662	236	-90	31.40	12.02	N	Y
14	22	464	417	252	-47	-212	10.21	45.69	Y	N
2617	18	60	141	205	81	145	135.06	241.31	N	N

Default MMRE = 71.27%
Calibrated MMRE = 84.55%
Default RMS = 210.34
Calibrated RMS = 157.67
Default RRMS = 75.77%
Calibrated RRMS = 56.80%
Default Pred (0.25) = 20.00%
Calibrated Pred (0.25) = 20.00%

Appendix E. Wilcoxon Signed-Rank Tests

Command and Control:

1. Uncalibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
9	517	1512	192	4	
141	322	1749	443	6	
145	101	200	98	2	
150	100	234	134	3	
155	74	91	23	1	
2517	167	579	247	5	
				21	0

Wilcoxon T = 21
n = 6

P	T_0 lower	T_0 upper	Result
0.062	1	20	Reject

Since Sum(Rank +) ≥ T_0 upper, SoftCost-R with default AWFFAC and KSLECPM overestimates.

2. Calibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
9	517	598	16	2	
141	322	735	128	6	
145	101	84	-17		3
150	100	98	-2		1
155	74	91	23	4	
2517	167	335	101	5	
				17	4

Wilcoxon T = 17
n = 6

P	T_0 lower	T_0 upper	Result
0.062	1	20	Accept

We cannot reject the hypothesis that SoftCost-R with calibrated AWFFAC and KSLECPM does not over or under estimate.

Signal Processing

1. Uncalibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
126	165	410	148	9	
130	738	615	-17		1
131	192	249	30	2	
134	228	381	67	6	
136	154	104	-32		3
137	274	515	88	8	
138	190	123	-35		4
144	145	255	76	7	
147	192	271	41	5	
				37	8

Wilcoxon T = 37
n = 9

P	T_0 lower	T_0 upper	Result
0.054	6	39	Accept

We cannot reject the hypothesis that SoftCost-R with default AWFFAC and KSLECPM does not over or under estimate.

2. Calibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
126	165	221	34	6	
130	738	331	-55		7
131	192	134	-30		5
134	228	205	-10		3
136	154	56	-64		8
137	274	278	1	1	
138	190	66	-65		9
144	145	137	-6		2
147	192	146	-24		4
				7	38

Wilcoxon T = 7
n = 9

P	T_0 lower	T_0 upper	Result
0.062	6	39	Accept

We cannot reject the hypothesis that SoftCost-R with calibrated AWFFAC and KSLECPM does not over or under estimate.

Unmanned Space - U.S.

1. Uncalibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
86	200	95	-53		3
112	1511	132	-91		5
113	1248	310	-75		4
306	94	106	13	1	
2622	558	293	-47		2
				1	14

Wilcoxon T = 1
n = 5

P	T_0 lower	T_0 upper	Result
0.062	0	15	Accept

We cannot reject the hypothesis that SoftCost-R with default AWFFAC and KSLECPM does not over or under estimate.

2. Calibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
86	200	161	-20		1
112	1511	223	-85		5
113	1248	523	-58		4
306	94	52	-45		3
2622	558	378	-32		2
				0	15

Wilcoxon T = 0
n = 5

P	T_0 lower	T_0 upper	Result
0.062	0	15	Reject

Since Sum(Rank -) $\geq$ T_0 upper, SoftCost-R with calibrated AWFFAC and KSLECPM underestimates.

Unmanned Space - European

1. Uncalibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
2574	181	527	191	5	
2575	25	139	456	7	
2577	768	341	-56		3
2579	410	496	21	1	
2581	764	465	-39		2
2598	45	217	382	6	
2607	37	77	108	4	
				23	5

Wilcoxon T = 23
n = 7

P	T_0 lower	T_0 upper	Result
0.046	2	26	Accept

We cannot reject the hypothesis that SoftCost-R with default AWFFAC and KSLECPM does not over or under estimate.

2. Calibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
2574	181	397	119	5	
2575	25	105	320	7	
2577	768	257	-67		4
2579	410	374	-9		1
2581	764	351	-54		2
2598	45	164	264	6	
2607	37	58	57	3	
				21	7

Wilcoxon T = 21
n = 7

P	T_0 lower	T_0 upper	Result
0.046	2	26	Accept

We cannot reject the hypothesis that SoftCost-R with calibrated AWFFAC and KSLECPM does not over or under estimate.

Ground in Support of Space - U.S.

1. Uncalibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
74	80	134	68		6
79	432	575	33	3	
88	244	1220	400	12	
89	602	2347	290	10	
90	1055	991	-6		1
93	401	2608	550	13	
97	530	834	57	4	
98	86	942	995	15	
99	234	83	-65		5
106	206	170	-17		2
114	235	1700	623	14	
116	1468	4168	184	7	
118	765	3734	388	11	
119	787	2905	269	9	
331	18	58	222	8	
				106	14

Wilcoxon T = 106
n = 15

P	T_0 lower	T_0 upper	Result
0.048	25	95	Reject

Since Sum(Rank +) ≥ T_0 upper, SoftCost-R with default AWFFAC and KSLECPM overestimates.

2. Calibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
74	80	99	24	3	
79	432	426	-1		1
88	244	875	259	12	
89	602	1683	180	10	
90	1055	711	-33		4
93	401	1870	366	13	
97	530	598	13	2	
98	86	675	685	15	
99	234	60	-74		6
106	206	122	-41		5
114	235	1219	419	14	
116	1468	2989	104	7	
118	765	2678	250	11	
119	787	2083	165	9	
331	18	39	117	8	
				104	16

Wilcoxon T = 104
n = 15

P	T_0 lower	T_0 upper	Result
0.048	25	95	Reject

Since Sum(Rank +) $\geq$ T_0 upper, SoftCost-R with calibrated AWFFAC and KSLECPM overestimates.

Ground in Support of Space - European

1. Uncalibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
2531	345	1328	285	15	
2540	74	184	149	8	
2543	105	112	7	1	
2545	42	194	362	16	
2546	85	429	405	17	
2547	100	1022	922	25	
2550	89	245	175	10	
2551	65	194	198	14	
2554	31	245	690	21	
2555	103	848	723	23	
2556	12	112	833	24	
2558	292	562	92	6	
2561	331	480	45	4	
2562	234	296	26	3	
2563	196	174	-11		2
2584	12	72	500	20	
2586	186	1022	449	19	
2588	128	358	180	12	
2590	59	163	176	11	
2591	55	102	85	5	
2594	156	460	195	13	
2595	58	143	147	7	
2603	225	562	150	9	
2608	71	562	692	22	
2609	60	306	410	18	
				323	2

Wilcoxon T = 323
n = 25

z = 4.318559

P	z_alpha/2	Result
0.05	1.96	Reject

Since z ≥ z_alpha/2 and Rank + > Rank -, SoftCost-R with default AWFFAC and KSLECPM overestimates.

2. Calibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
2531	345	422	22.3	8	
2540	74	58	-21.6		7
2543	105	36	-66		17
2545	42	62	48	12	
2546	85	136	60.0	15	
2547	100	325	225	25	
2550	89	78	-12.4		5
2551	65	62	-5		1
2554	31	78	151.6	21	
2555	103	269	161	23	
2556	12	36	200	24	
2558	292	179	-39		10
2561	331	153	-54		13
2562	234	94	-59.8		14
2563	196	55	-72		18
2584	12	23	92	20	
2586	186	325	75	19	
2588	128	114	-11		3
2590	59	52	-11.9		4
2591	55	32	-42		11
2594	156	146	-6		2
2595	58	45	-22.4		9
2603	225	179	-20		6
2608	71	179	152.1	22	
2609	60	97	62	16	
				205	120

Wilcoxon T = 205
n = 25

z = 1.143544

P	z_alpha/2	Result
0.05	1.96	Accept

We cannot reject the hypothesis that SoftCost-R with calibrated AWFFAC and KSLECPM does not over or under estimate.

Military Mobile

1. Uncalibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
348	396	320	-19		1
2456	221	630	185	5	
2502	633	375	-41		4
2503	783	476	-39		3
2507	647	434	-33		2
				5	10

Wilcoxon T = 5
n = 5

P	T_0 lower	T_0 upper	Result
0.062	0	15	Accept

We cannot reject the hypothesis that SoftCost-R with default AWFFAC and KSLECPM does not over or under estimate.

2. Calibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
348	396	790	99	5	
2456	221	343	55	4	
2502	633	529	-16		2
2503	783	718	-8		1
2507	647	843	30	3	
				12	3

Wilcoxon T = 12
n = 5

P	T_0 lower	T_0 upper	Result
0.062	0	15	Accept

We cannot reject the hypothesis that SoftCost-R with calibrated AWFFAC and KSLECPM does not over or under estimate.

Avionics

1. Uncalibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
11	198	539	172	4	
12	112	313	179	5	
13	752	988	31	2	
14	464	417	-10		1
2617	60	141	135	3	
				14	1

Wilcoxon T = 14
n = 5

P	T_0 lower	T_0 upper	Result
0.062	0	15	Accept

We cannot reject the hypothesis that SoftCost-R with default AWFFAC and KSLECPM does not over or under estimate.

2. Calibrated validation points:

Record #	Actual (PM)	Predicted (PM)	% Difference	Rank +	Rank -
11	198	374	89	3	
12	112	251	124	4	
13	752	662	-12		1
14	464	252	-46		2
2617	60	205	242	5	
				12	3

Wilcoxon T = 12
n = 5

P	T_0 lower	T_0 upper	Result
0.062	0	15	Accept

We cannot reject the hypothesis that SoftCost-R with calibrated AWFFAC and KSLECPM does not over or under estimate.

References

Boehm, Barry W. Software Engineering Economics. Englewood Cliffs NJ: Prentice-Hall, 1981.

Brooks, Frederick P., Jr. The Mythical Man-Month. Reading MA: Addison-Wesley Publishing Company, 1975.

Bruscino, Joe. Software Support Programmer for SoftCost-R. Telephone interviews about SoftCost-R source code. August 1996.

Collins, Anthony J. President, Resource Calculations, Inc., Englewood CO. Electronic mail correspondence. 23 August 1996.

Conte, Samuel D., H.E. Dunsmore, and V.Y. Shen. Software Engineering Metrics and Models. Menlo Park CA: Benjamin/Cummings, 1986.

Ferens, Daniel V. Class Handouts, Cost 677 Note-Taking Devices. School of Logistics and Acquisition Management, Air Force Institute of Technology, Wright-Patterson AFB OH, Fall 1995.

Ferens, Daniel V. and David S. Christensen. Software Cost Model Calibration - An Air Force Case Study. Air Force Institute of Technology, 1995.

Galonsky, James C. Calibration of the PRICE S Software Cost Model. MS Thesis, AFIT/GCA/LAS/95S-1. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A301377).

Jones, Capers. Assessment and Control of Software Risks. Englewood Cliffs NJ: Prentice-Hall, 1994.

Kressin, Robert K. Calibration of SLIM to the Air Force Space and Missile Systems Center Software Database. MS Thesis, AFIT/GCA/LAS/95S-6. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A301603).

Lieblein, Edward. "The Department of Defense Software Initiative," Communications of the ACM, 29(8):734-743, August 1986.

Management and Consulting Research, Inc. and Cost Management Systems, Inc. SMC SWDB Space and Missile Systems Center Software Database User's Manual Version 2.1. El Segundo CA: SMC/FMC, 1995.

Marsh, Alton. "Pentagon Up Against a Software Wall," Government Executive, 62-63, May 1990.

Ourada, Gerald L. Software Cost Estimating Models: A Calibration, Validation and Comparison. MS Thesis, AFIT/GSS/LSY/91D-11. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1991 (AD-A246677).

Pacheco, Thomas. An Investigative Search of Variables Impacting Software Support Costs. MS Thesis, AFIT/GIR/LAS/87D-4. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1987 (AD-A188879).

Putnam, Lawrence H. and Ware Myers. Measures for Excellence: Reliable Software on Time, Within Budget. Englewood Cliffs NJ: Prentice-Hall, 1992.

Rathmann, Kolin D. Calibration and Evaluation of SEER-SEM for the Air Force Space and Missile Systems Center. MS Thesis, AFIT/GCA/LAS/95S-9. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A300703).

Reifer Consultants, Inc. SoftCost-R MS/DOS Software Version 8.0 Manual. Torrance CA: Reifer Consultants, Inc., 1989.

Resource Calculations, Inc. IEC's Calibration Method for SoftCost-R Productivity Parameters. Englewood CO: Resource Calculations, Inc., Undated.

SMC SWDB. Version 2.1, IBM, five 1.4 Mb disks. Computer software. SMC/FMC, El Segundo CA, 1995.

SoftCost-R. Version 8.4, IBM, two 1.4 Mb disks. Computer software. Resource Calculations, Inc., Englewood CO, 1994.

Stukes, Sherry. Space and Missile Systems Center Software Database Collection Effort Final Report. Contract F04701-95-D-0003, Task 003. Oxnard CA: Management Consulting and Research, 15 November 1995.

-----. "Space and Missile Systems Center Software Database." Presentation to Professor Daniel Ferens, Captain Karen Mertes, and Captain Steven Southwell, Air Force Institute of Technology (AFIT/LAS), Wright-Patterson AFB OH. 2 April 1996.

Taub, Audrey E. "Calibration of Software Cost Models to DOD Acquisitions," Analytical Methods in Software Engineering Economics, First Annual Conference. 171-191. New York NY: Springer-Verlag, 1993.

Thibodeau, Robert. An Evaluation of Software Cost Estimating Models. Rome Air Development Center, Air Force Systems Command, Griffiss AFB NY, September 1991 (AD-A104226).

Vegas, Carl D. Calibration of Software Architecture Sizing and Estimation Tool. MS Thesis, AFIT/GCA/LAS/95S-1. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A301376).

Weber, Betty G. A Calibration of the REVIC Software Cost Estimating Model. MS Thesis, AFIT/GCA/LAS/95S-1. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A300694).

Wellman, Frank. Software Costing: An Objective Approach to Estimating and Controlling the Cost of Computer Software. New York NY: Prentice-Hall, 1992.

Vita

Captain Steven V. Southwell is from Mt. Morris MI. He graduated from the University of Illinois at Urbana-Champaign in 1988 with a Bachelor of Science degree in Computer Engineering. After receiving his commission into the United States Air Force through the Officer Training School in 1989, Captain Southwell was assigned to the Aeronautical Systems Center (ASC), Engineering Directorate, Avionics Engineering Division, Computer Resources Branch (ENASC) at Wright-Patterson AFB OH.

During his tour at Wright-Patterson AFB, Captain Southwell served as a computer resources engineer in support of several programs in the Aircraft Systems Program Office (SPO), Bombers and Tankers Division (SDB) and Transports Division (SDC). These programs included the KC-135 Avionics Modernization Program (AMP), the C-29A Combat Flight Inspection (CFIN) aircraft, the C-18D Cruise Missile Mission Control Aircraft (CMMCA), the C-135 Transport Advanced Avionics Cockpit Enhancement (TRAACE) program, and the Transport/Tanker Trainer System (TTTS) and C-27 request for proposal (RFP) efforts. In 1992, he was assigned to the National Air Intelligence Center (NAIC), Technical Assessments Directorate, Space and Technical Research Division, Engineering Branch (TAPE), Wright-Patterson AFB, OH, where he served three years as a space electronics engineer working in special projects. Captain Southwell entered the Air Force Institute of Technology at Wright-Patterson AFB OH in 1995 and graduated in 1996 with a Masters degree in Systems Management. He was subsequently assigned to the Human Systems Center (HSC), Human Systems SPO, Information Systems Division (YAI), Brooks AFB TX.

Permanent Address: 9131 N. Genesee Rd.
Mt. Morris MI 48458

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Master's Thesis

4. TITLE AND SUBTITLE
CALIBRATION OF THE SOFTCOST-R SOFTWARE COST MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

5. FUNDING NUMBERS

6. AUTHOR(S)
Steven V. Southwell, Captain, USAF

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
2750 P Street
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GSM/LAS/96S-6

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

SMC/FMC
2430 E El Segundo Blvd #2010
El Segundo CA 90245-4687

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 Words*)

This research effort focused on the calibration of SoftCost-R Version 8.4 to specific stratified data sets obtained from the Space and Missile Systems Center (SMC) Software Database Version 2.1 (SWDB). The accuracy of the new calibrated inputs was verified through comparisons between the calibrated and default estimates and the actual cost data. Statistical methods used to make the comparisons included magnitude of relative error (MRE), mean magnitude of relative error (MMRE), root mean square (RMS), relative root mean square (RRMS), and prediction level Pred (k/n) or percentage of estimates within (100 * k/n)% of the actual costs. The new calibrated parameters resulted in more accurate effort estimates and the calibration method appeared to be valid. However, the accuracy improvement was neither complete nor all encompassing. That is, the calibrated goodness of fit did not meet Conte's criteria of MMRE $\leq$ 25%, RRMS $\leq$ 25%, or Pred (0.25) $\geq$ 75%, and not all of the data sets achieved significant accuracy improvement due to the calibration. This result agrees with previous studies and emphasizes the need for complete, accurate, and homogeneous data.

14. SUBJECT TERMS
Software, Calibration, Cost Estimating, Cost Model, Software Cost Model, Department of Defense

15. NUMBER OF PAGES
100

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	

AFIT Control Number AFIT/GSM/LAS/96S-6

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to determine the potential for current and future applications of AFIT thesis research. **Please return completed questionnaire** to: AIR FORCE INSTITUTE OF TECHNOLOGY/LAC, 2950 P STREET, WRIGHT-PATTERSON AFB OH 45433-7765. Your response is **important.** Thank you.

1. Did this research contribute to a current research project? a. Yes b. No

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not researched it?
a. Yes b. No

3. **Please estimate** what this research would have cost in terms of manpower and dollars if it had been accomplished under contract or if it had been done in-house.

Man Years_____ $_____

4. Whether or not you were able to establish an equivalent value for this research (in Question 3), what is your estimate of its significance?

a. Highly Significant b. Significant c. Slightly Significant d. Of No Significance

5. Comments (Please feel free to use a separate sheet for more detailed answers and include it with this form):

_____ _____
Name and Grade Organization

_____ _____
Position or Title Address